Semantic innovation and grammaticalization of the English loanword type in colloquial Hindi

https://eprints.bbk.ac.uk/id/eprint/50603/

Version: Full Version

Deposit Guide
Contact: email

Semantic innovation and grammaticalization of the English loanword *type* in colloquial Hindi

EMMA THERESA WALTERS

Submitted for the degree of

PhD in Applied Linguistics

BIRKBECK, UNIVERSITY OF LONDON

Acknowledgements

In the remote but beautiful town of Pithoragarh, as a language learner grappling with Hindi in its many forms, I was constantly bemused and amused to find familiar English words cropping up but with very different meanings and applications. I quickly understood that the words were, in fact, no longer English, but had taken on a life of their own in Hindi. This was my introduction to loanwords and lexical semantics, although I didn't know it at the time. Fast forward a few years to a wintry afternoon at Senate House library in London and I retrieve a small bound volume containing a paper by Kay (1997). This was my introduction to type noun grammaticalization, although I didn't know it at the time. But a single sentence in that short paper by Kay blew my mind: ''I have, in fact, found no evidence that the hedges *kind of* and *sort of* contain the nouns *kind* and *sort*.' (1997:146) What?! A noun that is no longer a noun?! Both of these revelations fuelled a fascination which continues to this day. Although the quest is far from over, I have satisfied at least some of my intellectual curiosity in writing this thesis.

I want to thank my supervisor, Penelope Gardner-Chloros, from my heart for your support and your enduring patience, and for sharing my joy at discovering how Hindi *type* functions. I also want to thank every person who has ever said an encouraging word to me. Writing a PhD is like running several marathons, and every time anyone said, "You can do it!" or "You've got this!" it gave me a boost of strength. It wouldn't have been possible without the friends, family, colleagues and research staff who have cheered me on throughout the journey in ways great and small.

Many of those friends are from Pithoragarh. बहुत-2 धन्यवाद to everyone who has supported me in completing this research: to Ankita, Asha and Vandana for painstaking transcription; to Shobha Di for being a sounding board while I tested early hypotheses about *jaisa* and for explaining various cultural references from the web data; to Deepa and Bhagwat for being prepared to answer random questions on any given day; to my research participants for your willingness and positivity over several stages of research; and to all the staff and students of the ILC who always welcome me "home". To all the residents of our beloved town, I hope that this thesis puts Pithoragarh on the map and gives voice to your way of speaking Hindi.

To my parents, whose attic has been both office and bedroom for the past four years, thank-you for your incredible generosity and for going the distance with me. With lots of love to

you and to my wonderful brother and sister-in-law, sister and brother-in-law, and two super nieces, who bring me great joy: *Bua/Taiji* has finished her 'book' now!

Finally, I dedicate this thesis to the memory of Suman Smith, a friend I lost while writing this thesis. Su would have been so happy to know it was completed, and I like to think she's smiling now.

Transliteration and glossing conventions

Hindi transliteration

CONSONANTS	
Hindi character	**Roman transliteration**
क	ka
ख	kha
ग	ga
घ	gha
च	cha
छ	chha
ज / ज़	ja / za
झ	jha
ट	ta
ठ	thha
ड	da
ढ	dha
ड़	da
ढ़	dha
ण	na
त	ta
थ	tha

CONSONANTS	
Hindi character	**Roman transliteration**
द	da
ध	dha
न	na
प	pa
फ / फ़	pha / fa
ब	ba
भ	bha
म	ma
य	ya
र	ra
ल	la
व	va or wa
श	sha
ष	sha
स	sa
ह	ha
क्ष	ksha

CONSONANTS	
Hindi character	**Roman transliteration**
त्र	tra
ज्ञ	gya

VOWELS	
Hindi character	**Roman transliteration**
अ	a
आ	aa or a
इ	i
ई	i or ee
उ	u
ऊ	u or oo
ऋ	ri
ए	e
ऐ	ai or ei
ओ	o
औ	au

Notes on Hindi transliteration:

1. Hindi characters normally written as a *bindu* (ं), including अं, ङ and ञ, are transliterated as *n(a)* or its allophone *m(a)* with the exception of postposition में, which is transliterated as *me* following a local convention.
2. Hindi character अः is transliterated as *h*.
3. The letter *a* (and nasalized *an*) at the end of a word is always a long vowel sound (representing आ and आँ in Hindi).
4. Hindi character ऐ is consistently transliterated as *ai* with the exception of the first person singular pronoun मैं 'I' which is written as *mein* following a local convention and to avoid confusion with the common loanword *main* 'main'.
5. Transliteration follows dictionary spelling quite closely, except in case of ज़, which is rendered ज in the local dialect and pronounced as such by almost all speakers.

   Additionally, both the distal third person pronouns (वह/वे) are pronounced the same, as वो *vo*, and the Devanagari and transliteration both reflect this. Likewise with the

proximal third person pronouns यह/ये which are both pronounced ये and transliterated as *ye.*

Glossing

Leipzig Glossing Rules have been followed throughout the thesis. In addition to the standard list, the following abbreviations also appear:

ADV   adverb(-ial/-ializer)

EMPH   Emphatic

FILL   Filler

INTF   Intensifier

SD   Similative demonstrative

IPA

Phonetic representation in this thesis is according to the latest IPA chart produced by the International Phonetic Association, as located here:

https://www.internationalphoneticassociation.org/IPAcharts/IPA_chart_orig/IPA_charts_E.html, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=92652779

Table of Contents

List of Tables

List of Bar Charts

Abstract

Recent years have seen increased interest in the grammaticalization of type nouns cross-linguistically. This thesis contributes to the growing body of research in the field by using naturally occurring data to investigate innovative uses of the English loanword *type* in colloquial Hindi by residents of the North Indian Himalayan town of Pithoragarh. A semantic-pragmatic and syntactic analysis of the spoken data reveals that the functions of *type* in Hindi mirror the grammaticalized functions of *type* cognates in other languages, including uses as a similative and a hedge, and incipient uses as a quotative marker, a general extender and a discourse marker.

The emergence of *type* in a language already replete with lexical options offers a unique opportunity for comparison between the functions of *type* and those of equivalent Hindi lexemes. Of particular import is the analysis of *aisa* 'like this' and related lexemes, a class known as 'similative demonstratives' (van der Auwera and Sahoo, 2020). Highly prolific in Hindi, similative demonstratives share many of the extended functions of synonyms of *like* and of *type*, but also 'discourse-structuring devices' associated with English *such* and German *so* (König, 2020). A 'phoric' construction formed with loanword *type* appears in similar contexts to similative demonstrative *aisa* beyond the remit of a modifier, showing evidence of semantic-pragmatic extension and decategorialization. This thesis presents a motivated argument that *type* may have modelled itself on *aisa* and be grammaticalizing along a similar path or indeed paths.

Finally, the study offers a novel contribution to research into type noun grammaticalization due to *type*'s status as a recent loanword. Whilst being fully integrated into the Hindi lexicon there is evidence that loanword *type* also retains its status as an English-origin lexeme, functioning as a flag for English code-switches alongside its hedging role.

1   Introduction

This study focuses on a mountain community in North India's Hindi belt, the semi-rural town of Pithoragarh, representative of the myriad of multilingual speech communities in which English-medium education has become increasingly popular and yet lags behind the level of elite schooling available in North India's urban centres. Prolonged language contact between Hindi and English, which has become more pronounced in recent decades, has resulted in extensive lexical borrowing, with a large number of English loanwords entering the Hindi lexicon. As a long-term resident of Pithoragarh, the researcher's interest was piqued by the large number of loanwords being used with different meanings and to different pragmatic purposes than their semantic-pragmatic functions in English. This study seeks to provide answers to the question of how loanwords can come to acquire such different functions in the borrowing language.

The locus of the study is the everyday language of Pithoragarh residents. Recordings of informal conversations between Hindi speakers were conducted which constitute the first corpus of naturally occurring Hindi discourse collected in North India.[1] An initial survey of all English-origin words in the corpus revealed a large volume of loanwords appearing in varying frequencies and with varying levels of semantic shift. Loanword *type* (n.) was one of the most prevalent English lexical items to appear in the data. The first major goal of this research is to provide a comprehensive description of the functions of *type* in synchronic data, detailing both the semantic-pragmatic contexts in which it is used, and the syntactic constructions in which it occurs. Note that throughout this thesis, *type* in italics refers to the word, and the non-italicized compound 'type noun' refers to the group of nouns which traditionally fulfil a categorizer function, of which *type* is one example.

Cross-linguistically, cognates and synonyms of *type* have garnered interest in recent years as these so-called type nouns have been shown to grammaticalize and develop extensive pragmatic functions as phoric modifiers, similatives, quotatives, hedges and focus markers, among other functions.[2] A further goal of the study is to investigate how these patterns of usage compare with *type* cognates in other languages: is Hindi *type* following a cross-linguistic precedent? A cross-linguistic analysis based on studies of synchronic and diachronic

---

[1] This statement is true at the time of writing. Note that the broad definition of 'corpus' is used here denoting a logical and organised collection, in this case of audio recordings. A corpus of spoken Hindi recordings from the TV (Parshad et al., 2016) is, to the best of my knowledge, the only other corpus of spoken Hindi discourse that has been created.
[2] Here and throughout, 'phoric' is used as a cover term subsuming both anaphora and cataphora.

uses of *type* cognates in other languages is used to explore this question. A semantic-pragmatic and syntactic analysis of the functions of loanword *type* in the colloquial Hindi from Pithoragarh reveals that the innovative functions of *type* are consistent with those of its cognates around the world, although some important differences are highlighted.

In order to situate the forms and functions of *type* within its broader context, the thesis undertakes the somewhat daunting task of documenting and describing several semantic-pragmatic aspects of Hindi language use which have hitherto been underrepresented in the literature. Daunting due to the sheer scale of the endeavour, the thesis analyses several of the extended uses of similative demonstrative *aisa* and its related lexemes (notably relative adjective *jaisa*) and compares their usage to that of *type*: this is the third goal of the study. Findings show that *aisa* and its related lexemes fulfil the vast array of pragmatic and discourse-structuring functions which have been developed by similative demonstratives cross-linguistically as well as by type nouns, including Hindi *type*.[3]

An explanation of the striking similarity between Hindi *type* and *aisa* is explored using a theoretical framework which combines the theory of grammaticalization (Hopper and Traugott, 2003; Heine and Kuteva, 2005) with the principle of 'attraction' that operates between lexemes within a lexicon (De Smet et al., 2018). This framework facilitates the fourth goal of the thesis, to consider whether the synchronic functions of *type* show indications of grammaticalization and if so, how this might be explained. The thesis concludes that *aisa* is functioning as a grammaticalization model for newcomer *type* and causing it to develop in innovative ways not observed in all other languages.

The text is arranged as follows. Following this introductory chapter, Chapters Two to Six set the scene and explain the background to the study, covering the research setting, the theoretical framework, and the methodology. **Chapter Two** provides background on language contact between Hindi and English in India, both historical and modern. **Chapter Three** situates the research project in its geographic and sociolinguistic context, commenting on each of the themes raised in Chapter Two. **Chapter Four** introduces lexical borrowing as an outcome of language contact and discusses the language-internal processes that are triggered by the introduction of a new lexeme into the lexicon. Specifically, the borrowing of a new lexeme into a semantic field engenders competition between the lexemes and can

---

[3] The lexeme *type* will be referred to variously throughout the thesis as both '(English) loanword *type*' and 'Hindi *type*'. The latter reflects its identity as an established lexeme in the Hindi lexicon, and the former reflects its origins.

result in attraction as well as functional specialization. **Chapter Five** introduces the phenomenon of grammaticalization and its theoretical framework and presents the English loanword *type* as having already undergone a process of grammaticalization in English. Having presented the background to the key research themes as well as the physical context of the study, in **Chapter Six** the methodology employed in answering the research questions will be outlined and an overview of the data presented.

Combining a review of relevant literature with analysis of its contents, **Chapter Seven** presents the global picture by synthesizing research on type nouns cross-linguistically with a particular focus on cognates of *type*. The chapter demonstrates the high degree of semantic-pragmatic uniformity in the functions of *type* cognates in the world's languages, while also highlighting variation in their more subtle uses as well as their syntactic realization. Likewise, analysis of the grammaticalization of a class of words known as 'similative demonstratives' (Sahoo and van der Auwera, 2019) cross-linguistically shows a great degree of overlap between their extended functions and those of grammaticalized *type*. This is highly pertinent to Hindi since similative demonstratives are prolific in Hindi, particularly in colloquial discourse. The chapter ends with a brief review of how similative demonstratives and type nouns interact in those languages for which published research on both is available.

Chapters Eight and Nine are the core of the thesis. In this crucial **Chapter Eight**, each of the semantic-pragmatic functions of *type* attested in the Hindi data are examined in turn, following a presentation of how the functions are fulfilled in Hindi by other equivalent lexemes. Loanword *type* is seen to mirror many of the innovative functions of Hindi similative demonstrative *aisa* and its related lexemes *aise*, *jaisa, jaise, vaisa* and *vaise*. **Chapter Nine** integrates the findings of the previous chapters, both the cross-linguistic evidence for the grammaticalization of *type* cognates and the analysis of the synchronic uses of *type*, into a motivated argument for the grammaticalization of *type* in Hindi, explained by the principle of attraction.

Chapters Ten and Eleven present the implications of the findings. These final chapters draw together the cross-linguistic analysis of the functions of *type* with the findings from the empirical study of *type* in (Pithoragarh) Hindi. **Chapter Ten** 'Discussion' highlights the similarities and differences between the innovative semantic-pragmatic functions fulfilled by Hindi *type* and those of other *type* cognates cross-linguistically. It also discusses the impact of similative demonstrative *aisa* on *type* in relation to similar (non-)occurrences in other languages, notably in the context of Norwegian where a similar phenomenon is observable.

In addition, the chapter contributes to the field by providing metacommentary on the difficulties encountered in the cross-linguistic comparison of type nouns. **Chapter Eleven** 'Conclusion' summarizes the findings of the thesis and focuses on its contributions to our understanding of loanword grammaticalization, the interrelationship between similative demonstratives and type nouns in general, and innovative semantic-pragmatic functions of Hindi lexemes in particular.

2 Language contact between English and Hindi in India

This chapter considers previous research most pertinent to the topic of study, starting with a critical look at the status of English in India and different types of Hindi–English language mixing described in the literature. The second part of the chapter consists of an overview of theories and research relating to language mixing cross-linguistically. Lexical borrowing and code-switching are discussed, the latter of which has been afforded considerable academic attention in the last thirty years. Finally, I consider the issue of agency in language change and conclude that both the individual, their social network and the broader societal context play a part in influencing which lexemes will be appropriated into a language and in what ways.

2.1 Historical contact

The history of English in India began with the arrival of the East India Company (EIC) in the early 1600s. Although English was eventually to become the colonial language of the British Empire and to be used for all bureaucratic and legal purposes, the process started with the learning of English by local merchants and rulers who had a vested interest in maintaining relations with the EIC (Krishnaswamy and Burde, 1998). In the early 1830s, the British government restricted the activities of the EIC, which was subsequently unwilling to finance the 'importation' of British writers to complete administrative work. The creation of public sector jobs that required English skills was thus a financial motivation for Indians to learn English. In order to train the new elite class of Indian clerks, English university education was established, and that precipitated the subsequent instigation of English-medium schools at secondary and primary levels (Krishnaswamy and Burde, 1998; Annamalai, 2001). Graddol (2010) argues that the decision to produce an English-educated class of Indian graduates was primarily motivated by financial constraints and not a desire to produce a linguistic hegemony. That the motivation for acquiring English has always been for economic gain is of significance for our understanding of modern India. English is still (legitimately) associated with wealth, as a study of high school students by Vaish (2008) study shows and as many of the participants of this study attest to.

Despite the necessity of English for Indian civil servants working alongside the British administration, these elite workers remained a tiny percentage of the overall population of India. The official census report from the census of 1921 lists the total population of India at almost 319 million people, and the same report states the number of English speakers to be

308,071; these people are purported to be predominantly British but also including a small number of Americans, Canadians and Australians.[4] This means that there were fewer than one English speaker to 100,000 Indians in the country, which is not a viable statistic for the effective transmission of a language. Although a proportion of Indians were trained and well-versed in English during the British Raj, they were a very small minority.

When India gained independence from England in 1947, English was included in the Constitution as an interim official language. It was intended to be phased out by 1965, by which time Hindi would have been sufficiently developed on a national level to replace it. However, various political and economic conditions prevented this from happening, the most pressing of which was a series of language riots in the state of Tamil Nadu, which compelled President J. Nehru to amend the Constitution and retain English as an official language alongside Hindi (Vaish, 2008). English was thus never replaced as the language of the courts and central government administration but remained a language of the elite. It was the development of an open market economy in the 1990s and the rise of English as a global language that precipitated its expansion throughout the social strata.

2.2   Multilingual India

Indian society and culture are renowned for multiplicity and variety, and the Indian linguistic context is no less diverse. The Ethnologue website lists 415 living languages spoken in India;[5] monolingualism is rare if not non-existent, multilingualism the norm. Bhatia (2011) provides insight into his own language practices speaking four languages with his extended family at home; particularly notable is the way in which his choice of language is determined by multiple factors including interlocutor, subject matter and emotional state at the time of speaking. Knowledge and use of English is considered a necessity, a theme which appears in Vaish (2008)'s interviews with secondary school students from a socio-economically disadvantaged part of Delhi and in conversations with research participants from Pithoragarh. As a consequence of diglossia, English is the language of the courts and of higher education and is also essential for passing competitive exams and acquiring well-paid jobs. However, in multilingual India, elite L1 speakers of Indian English also need Hindi to speak to lower status workers or home helps (Chand, 2011).

---

[4] All data was contained in a digital (PDF) version of the document 'Census of India, 1921. Volume I. Part II – TABLES' accessed from https://censusindia.gov.in/DigitalLibrary/browseyearwise.aspx on 11th August, 2020. Document URL:
https://censusindia.gov.in/DigitalLibrary/data/Census_1921/Publication/India/24161_1921_TAB.pdf
[5] Accessed via Wikipedia – this number was 447 when last accessed via Ethnologue in 2016.

The number of Indians who currently speak English is a matter of debate and the answer will of course differ depending on the definition of 'speak English' employed. Graddol (2010:66) cited a number of statistical sources, each of which approached the question of how many Indians speak English from a different angle, and concluded: '[n]o-one really knows how many Indians speak English today – estimates vary between 55 million and 350 million – between 1% of the population and a third.' However, according to the 2011 census, only 259,678 people listed English as their "mother tongue".[6] Compared to the number of residents who selected Hindi as their mother tongue, numbering over 322 million, the number of L1 English speakers in the nation is miniscule. And yet, the census results indicate that there are millions of Indians who have English as a second or third language.[7] The impact of English on the Hindi (or other first language) of Indian speakers will be determined by the level of contact, which will be different for different strata in society.

2.3   Modern contact

The economic boom of the 1990s was the stimulus for an increased interest in English by the masses because opening up the nation to foreign investment facilitated whole new service industries in IT development and other forms of business process outsourcing. These new jobs were more highly dependent on English skills as the client base was exclusively in the English-speaking nations. New opportunities brought concomitant competition for jobs and as more and more people rejected their parents' agricultural occupation in search of a higher and more stable income. According to Parshad et al., '[j]obs in almost all prestigious sectors require English proficiency' (2016:378). Government jobs remain a holy grail for many families as the associated pension schemes ensure financial provision into old age. The application process for all government jobs involves sitting the relevant highly competitive exam. The exam papers are sometimes available in both Hindi and English, but applicants often prefer to respond in English as the questions are often phrased in a manner that is easier to comprehend in English than Hindi.[8] Thus English language capability as defined through the ability to 'crack' competitive exams became the gateway to accessing these most sought-after jobs and led to the explosion of private English-medium education, the English coaching centre industry and the competitive exam coaching industry. English was the

---

[6] This figure includes responses from any foreigners who were resident in India at the time of the census, which incidentally included the researcher! All statistics have been taken from the official census website https://censusindia.gov.in/2011census/. Last accessed 12th August, 2020.
[7] For example, there were 32 million who claimed to have Hindi as a first language and English as a second language.
[8] Personal communication with the daughter of a research participant who spent seven years studying for competitive exams before finally getting a government job.

language the lower and more disadvantaged classes always aspired to learn, a feat now made possible with the advent of reasonably priced English-medium schools and further encouraged through film and TV, and the explosion of social media such as Facebook and WhatsApp, which are extremely prevalent in India.

2.3.1 English in the Indian education system

The rise of English-medium education in India is therefore the mechanism by which English has proliferated, to differing degrees in different locations (urban and rural), with access to this education almost entirely dependent on socio-economic status. There is a dichotomy in India between what has traditionally been state-funded vernacular-medium schools and private English-medium schools. This dichotomy is both perceived and actual: perceived because there are students from certain Hindi-medium schools who speak better English than some students from nominal English-medium schools; actual, because students from English-medium schools tend to have better English than students from Hindi-medium schools.[9] LaDousa's (2014) book *Hindi is our ground, English is our sky* is a study of values and practices concerning language and education in the city of Varanasi, North India. LaDousa presents the languages Hindi and English as diametrically opposed markets in which an English-medium education is seen as intrinsically more valuable than a Hindi-medium education. The hierarchy extends even further such that certain types of English-medium education are preferred. For the higher classes, high fee-paying schools within the city are considered to offer a better standard of (English) education than lower quality local English-medium schools. For the elite, whose children study at boarding schools of national reputation, the English of value is that which offers international scope.

Most Indians learn English as a school language. A very limited percentage of the population, referred to as the "elite" in the literature (Chand, 2011; Mohan, 2014), learn English as an L1 and use it at home. Students who attend an English-medium school start acquiring English as young as two years old with their entry into Nursery class at English-medium schools. Their peers who attend Hindi-medium schools currently start learning English immediately, whereas many of their parents did not start until 6th standard (age 10).

The Three-Language Formula (TLF), a linguistic educational policy introduced in 1968, was designed to ensure that linguistic bridges were built between Hindi-dominated Indo-European North and Dravidian South India. However, in practice, the implementation of the

---

[9] This statement is based on experiences of Hindi-medium and English-medium students joining the International Learning Centre (ILC) over a six-year period, 2010-2016.

TFL has ensured the increased promulgation of English throughout the county, but not the north-south linguistic harmony that had been conceptualized, since other languages spoken regionally are preferred as the third language.

2.3.2 Technology and social media

The open market economy introduced into India in the 1990s brought sweeping changes in technology and social media. From the technological point of view, a rapid increase in

Leading countries based on Facebook audience size as of July 2020 (in millions)
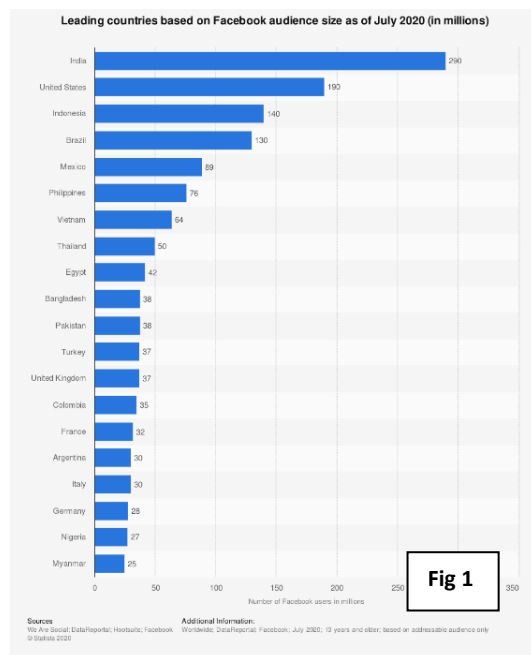
production of mobile phones (see Jeffrey and Doron, 2013) and access to social media through the internet have brought millions of Indians into greater contact with English and with the Roman script than ever before. As of July 2020, India has the largest Facebook population globally, with over 290 million users (see Fig 1).[10] Most of these access Facebook and other social media platforms like WhatsApp via mobile phones using internet provided through mobile networks. Facebook launched its mobile Facebook capability in 2007 and its Android app in 2010.

Computers and mobile phones were released with only the Roman alphabet on the keyboard. The legacy of phone manufacturers and software companies promoting their products in India before Hindi menus were made available is that users had to learn to use the devices in English and type in Roman script, and this has become standard throughout North India.

2.3.3 Film and TV

Si (2011) analysed the scripts of seven Bollywood films spanning three decades from approximately 1980, 1990 and 2000 and compared the mix of Hindi (or Punjabi) and English found in the speech of adults and young people. Despite the data being fictional, it afforded the researcher the opportunity to carry out a longitudinal study. Si (2011) found an overall increase in the usage of English, characterized by insertions of English words into Hindi matrix sentences and alternations between Hindi phrases and English phrases (or vice versa). The

---

[10] https://www.statista.com/statistics/268136/top-15-countries-based-on-number-of-facebook-users/ Accessed on 13th August 2020

author further argued that this was a reflection of changes in the linguistic practices of the urban elite (upper-middle-class) in India. The researcher attributes the changes in large part to the availability of technologies like mobile phones, satellite TV and the internet. There is pervasive usage of English lexemes in Indian media, but to differing levels, depending on both the channel and the target audience of the programme. Among the most popular forms of media are Bollywood films, which increasingly incorporate Hindi-English code-switching or Hinglish. A popular comedy show called The Kapil Sharma Talk Show, named after its comedic host, started during the period of data collection and is an example of the kind of TV programme which one needs a certain level of English to fully understand. However, the level of code-switching is likely to be managed in order to allow the maximum audience nationwide (see Si (2011:403) for a similar claim related to code-switching in Bollywood films) and the jokes are primarily Hindi-based and not dependent on an understanding of English.

Having outlined the history of language contact between Hindi and English in India as a whole as well as the contemporary situation, the following chapter (Chapter 3) focuses on language contact in the research site for this study, the Himalayan town of Pithoragarh.

3   Language contact between English and Hindi in Pithoragarh

In this chapter, the reader is introduced to the site of data collection for this thesis: the semi-rural town of Pithoragarh in North India. The town is presented from the perspective of its remote geographical location and significance as a border area. Then the multilingual context of Pithoragarh society is described, leading to the second half of the chapter, which is focused on the history of English influence in Pithoragarh, an evaluation of both historical and modern influences and the resulting impact on the local languages. Despite the early influence of army connections and missionary presence it is concluded that the rise in English-medium education would have had the biggest influence on Hindi, supported by social media and, to a lesser degree, TV and film. Lexical borrowing into Hindi from English is one of the most significant outcomes of the longevity of contact in Pithoragarh.

3.1   Geographical and socio-political context

The town of Pithoragarh has a population of approximately 80,000 and is the capital of the largest district in the Kumaon Hills, a lower Himalayan mountain range which comprises half of the state of Uttarakhand and is the homeland of over 3 million Kumaonis. Uttarakhand is a predominantly Hindu region known as *dev bhoomi* 'land of the gods' and famous for being the source of the Ganges and the site of numerous temples and pilgrimage sites. Fig 2 is a map of Uttarakhand with the districts delineated and Pithoragarh town marked with a yellow star. The district of Pithoragarh boasts several mountain peaks above 6000m. There have been recent efforts to increase the attraction of the area as a tourist destination, including the renovation of an old eighteenth century fort (opened to the public in 2017).

Fig 2: A map of India with details of the state of Uttarakhand

The remoteness of the town's location (a seven-hour drive from the nearest railhead) means that tourists, either domestic or foreign, are few and far between. Pithoragarh's airport opened to light commercial aircraft in 2018 after more than twenty years of development but had to close soon after. At the time of writing, it services commercial helicopters. Although there have been improvements in road widening and the erection of safety barriers in recent years, there are still frequent deaths on Kumaon's dangerous mountain roads and landslides are common during the monsoon season, further increasing the danger. These combined factors mean that the district remains largely isolated. Pithoragarh's position on the border with both Tibet and Nepal, both with contentious border disputes, marks it out for special defence arrangements. The district is home to large brigade headquarters as well as bases for two border police organizations, the Indo-Tibetan Border Police (ITBP) and *Sashastra Seema Bal* (SSB).

Having introduced the physical setting of the research location of Pithoragarh, the next section describes the town's sociolinguistic composition.

3.2   Sociolinguistic context

3.2.1   Multilingual Pithoragarh

The previous chapter described the highly multilingual nature of Indian society and Pithoragarh is no exception to this. Annamalai outlines the division of functions between different languages in India as follows: 'One language is for ethnic identity, another for business transactions, another for official dealings, another for entertainment, another for rituals, and so on.' (2001:36) In Pithoragarh, this translates into <u>Kumaoni</u> for ethnic identity, <u>Hindi or English</u> for official dealings, a <u>Hindi-Urdu-Punjabi mix</u> for entertainment and a <u>Sanskrit-Hindi-Kumaoni mix</u> for worship and rituals. Pithoragarh residents either speak or have passive knowledge of at least two or, more commonly, three or four of the languages who usage is described briefly below:

- **Hindi** is the *lingua franca* in the town (see Appendix B for a summary of the non-standard features of the regional dialect of Hindi)
- **Sanskrit** is used in all major Hindu worship events and some knowledge is therefore essential
- **English** is necessary in order to get a well-paid job, but is not spoken fluently by many
- **Kumaoni** is the mother tongue of up to 50% of Pithoragarh residents, but does not enjoy the same prestige as Hindi
- Other languages spoken by residents of Pithoragarh who are not of Kumaoni ethnicity:
    - various **Tibeto-Burman languages** spoken by tribal people native to the region
    - **Nepali** and **Dotyali**, spoken by the population of Nepalese porters/labourers
    - **Punjabi** spoken by the large emigrant population from Punjab who own several of the shops in the town
    - **Bihari** spoken by more recent migrant labourers

Fig 3

In addition to being multilingual, Pithoragarh society is increasingly bi-graphic. A brief review of signboards in Pithoragarh town centre following a linguistic landscapes approach illustrates that they communicate using (a combination of) any of the following: Hindi words written in Devanagari, Hindi words written in Roman script, English words written in Devanagari and English words written in Roman script. Fig 3 is a photo of the researcher with former colleagues outside a local English language institute, the International Learning Centre (ILC), in 2018. The image shows the front of the ILC as well as *Pehli Manzil* 'First Stage',

another institute located on the floor above it. The signboards include signage painted onto the wall and provide examples of three of the four graphic combinations, including institute name *Pehli Manzil* which is transliterated into English but written in a pseudo-Devanagari style. The other combinations present are English written in Roman script and Hindi written in Devanagari.

Fig 4: Photo of a local bookshop taken by the author on 29th March, 2017

The photo in Fig 4, also from Pithoragarh town centre, illustrates a fourth option: English words transliterated into Devanagari. 'Mehta Book Depot' and 'Mehta Scientific' are the headings which appear above the shop front. Mehta is a common family surname and therefore the shop name contains a mix of Hindi and English words expressed in Devanagari, a common feature of local signs. A fifth possibility is included in Kathpalia and Wee Ong's (2015) analysis of Amul butter advertisement billboards, where part of a word is written in one script and the other part in another script, as in बutterर Vutteर (advertisement released by Amul in 2012 and reproduced in Kathpalia and Wee Ong, 2015:571); however, this is not a common feature of signs in Pithoragarh.

3.2.2 Pithoragarh as an educational hub

Pithoragarh has developed as an educational hub for the district and hosts more than 70 private schools, most of them English-medium. Socio-economic status marks the divide between students of government schools and those of private schools. The standard of English spoken in Pithoragarh is low compared to that of major urban centres like Delhi, but it has been steadily increasing in the past two decades. The Indian government's TLF policy of 1968, designed to foster unity among India's diverse peoples, requires each child to learn

three languages, two Indian and one foreign.[11] The interpretation of the government's TLF policy in Kumaon is such that all students in the region learn **Hindi** as a first and **English** as a second language (or the reverse order in English-medium schools), and **Sanskrit** as a third language (the only other language on offer). Despite the fact that Hindi is a second language for the majority of the rural population, schooling in the medium of their first language **Kumaoni** has never been orchestrated, since Kumaoni is not one of the 22 so-called 'scheduled languages' which are listed in Schedule 8 of the Indian Constitution. Nor is **Kumaoni** provided as a subject on the school curriculum in either government or private schools.[12] Some work has been done in the last five years to develop educational resources in Kumaoni, and a couple of pamphlet-style books have been produced, but there are not as yet any materials designed to teach the language, or any primary level school curricula published in the Kumaoni language.

In addition to the large numbers of private schools in the town, there are also many institutes and coaching centres helping (mainly) high school and university age students learn computer skills, prepare for the many competitive exams which are the gateway to government jobs, and improve their English, which is also essential for passing competitive exams. The number of coaching institutes has been on the rise in the past 15 years. The researcher was a founding team member and (EFL) English teacher of one such institute, the ILC, in 2007. The ILC is an English language school situated on the busy road between the town centre and Pithoragarh's university campus. When it opened, it was the second institute in Pithoragarh to offer classes in spoken English and it attracted an initial 48 students. During the busy season in 2019, ILC teachers taught a total of 600 students a day.

3.2.3 Socio-economic factors

Of the 80,000 residents of Pithoragarh town, many are educational migrants from the hundreds of villages in the district. Rental accommodation is a source of income for local families who build properties with an extra storey and rent out small flats to the many village families who stay in the town during term-time to educate their children in private schools and return to the village during holidays.[13] There is almost no local industry due to the

---

[11] The New Education Policy released by the central Indian government reiterates the TLF and still allows for these three languages to remain dominant in schools in Uttrakhand. See https://www.ndtv.com/education/national-education-policy-2020-what-three-language-formula, last accessed 18th May 2021.

[12] Since starting this project, one private school in Pithoragarh has added Kumaoni to its timetable as a weekly subject at primary level.

[13] I lived in two such flats during my nine years in Pithoragarh. In both cases, the landlord family lived upstairs, and I lived on the ground floor.

impracticalities of transportation on mountain roads. Kumaoni villagers are traditionally subsistence farmers, but there is also a high percentage of out-migration where one or more family members work in a big city (often in the hospitality industry) and sends money home. Being the district capital, Pithoragarh houses the court and government offices; government jobs are highly prized and sought after, and the more senior positions require candidates to score highly in exams with English as the examination medium. Their scarcity compared to the population of graduate students means that young people can spend years in exam-based coaching institutes and has also encouraged the popularity of English institutes such as the ILC (see section 3.2.2). Kumaon also has a strong tradition of joining the Indian Army. The Kumaon Regiment, which traces its history back to 1813, is most one of the oldest and most highly decorated regiments in the army and an estimated 30% of families have at least one family member serving. Army jobs are highly sought after, as they offer job security and a pension.[14]

3.3 History and influence of English in Kumaon

In this section several avenues for the influence of English on Hindi and Kumaoni are considered and evaluated in terms of their possible level of impact. Historical influences are considered first, namely the long tradition of enlisting in the army, and the presence of English-speaking missionaries in the town. English-medium education in the town is proposed to be the greatest route of influence from English, with the growing use of social media platforms a second route. The results of a questionnaire support this theory.

3.3.1 The Indian Army

The Kumaon Regiment boasts nineteen battalions and there are army posts in several locations around Kumaon, as well as units serving in troubled spots around India. Contact with English through the army has resulted in a large number of specific English-origin loanwords such as *पलटन paltan* 'platoon' but undoubtedly have contributed to the establishment of more general words like *officer*, *duty* and *retire*, which are commonly used in conversations with and about servicemen. However, the language of command in the British Indian Army was historically Urdu, which all British officers were expected to learn (Marston, 2003). This means that soldiers were not required to learn English in order to follow their command and we can infer that the influence of English from the army was likely limited to loanwords and did not result in transmission of spoken English as a medium of

---

[14] https://www.bbc.co.uk/news/world-asia-india-61836637

communication. The language of command in Kumaon regiments is currently Hindi; only officers communicate in English.

3.3.2 Christian missionaries

There was a continuous presence of (mainly) ladies from the London Missionary Society and the American Methodist Episcopalian Mission in Pithoragarh between 1871 and 1931. Anne Budden, the missionary in charge of the work in Pithoragarh for many years, is reported to have been fluent in Kumaoni (Badley, 1931:380). It seems that all missionaries in North India would have learnt Hindi initially and those posted in Pithoragarh may have also acquired Kumaoni. The missionaries founded two English-medium schools which are still running today - one for girls and one for boys. LWS Girls' Intercollege was established in 1871 and Mission School (for boys) was established in 1872.[15] Also established were a home for mistreated women, another for boys, another for girls, and a church.

Fig 5

In 1885, Mary Reed, a young missionary from Ohio, America, moved to Pithoragarh to recuperate having become ill in another city. It was later discovered that she had contracted leprosy from an unknown source, and following her diagnosis in the USA, she elected to return to India and work at the leprosy colony in Chandak, a hilltop village 5km above Pithoragarh. She lived and worked at the colony, supported by local staff, until she died there in 1943. The photo in Fig 5 is of the cross located at the leprosy mission and showing the view over Pithoragarh town. Accounts of Mary Reed's life by Mackerchar (n.d.) and Jackson (1899) state that she dedicated time initially to 'language study', understood to be Hindi, but it is likely that because of her longevity in the area and degree of isolation she would have learnt Kumaoni too. During her years in Chandak, Mary Reed's disease stabilized and she became

---

[15] https://mypithoragarh.com/schools/l-w-s-girls-inter-college-pithoragarh/
https://oruschool.in/india/mission-inter-collage-school-bhatkot. Accessed on 6th August 2020.

able to interact with non-lepers, teaching several young men English grammar (Mackerchar, n.d.).

An evaluation of the extent of influence of the missionaries on local language use concludes that their influence was probably limited to the group of children and adults with whom they interacted. In the case of Mary Reed, this was mostly restricted to the leper community. Anne Budden and colleagues who taught at the schools would have interacted with more local people who would themselves have been capable of transmitting English words, but this would most likely have been limited to the small Christian community and not had much impact on the rest of society. Furthermore, the missionaries made an effort to learn Indian languages and would likely have conversed in Kumaoni as much as they could. In any case, the small proportion of fluent English speakers in the town combined with the generally low level of English comprehension and usage discovered by the researcher in 2007 suggest that as with the Indian Army any linguistic influence resulted in the appearance of select English-origin loanwords in the local language(s) and not in the general acquisition of English per se.

3.3.3   Official documents

A further influence not on English fluency but on the spreading of English lexemes is the use of English for official documents, both legal and governmental. However, local court proceedings are conducted in Hindi and official government documents (for example, the 'Control Order' of 2015) are produced in both languages.[16] Professionals like lawyers and government workers could be relied upon to complete any paperwork requiring English and therefore the degree of influence on the average Pithoragarh resident is understood to have been minimal and limited to certain words.

3.3.4   English-medium education

LaDousa's (2014) description of language attitudes in Varanasi accurately reflects the dichotomy between English-medium and Hindi-medium education in Pithoragarh, and the region as a whole. In the case of Uttarakhand, government schools had always been Hindi-medium, until a very recent decision by the state government (2015) to convert certain government schools to English-medium. However, a very well-respected Hindi-medium school in the town has recently opened an English-medium track so parents can choose which medium they want their children to study through. Although there are no elite schools

---

[16] The Control Order details the rules related to government-subsidised food rations

in Pithoragarh, local residents are universally aware of which of the English-medium schools offer a better standard of English and those schools have correspondingly higher fees.

The schools opened by the missionaries at the end of the nineteenth century are understood to be the first English-medium schools in the town. In recent years, starting with the establishment of the first CBSE-affiliated high school in the town, this number has increased exponentially.[17] The rise of English-medium education with textbooks printed in English has greatly increased the level of contact with English experienced by all generations of Pithoragarh residents. Not only do children have to use solely English at school and complete their homework and exams in English, but parents and grandparents are also affected. Children studying in English-medium schools will develop what Matras (2009) terms 'domain specialisation', where their school-based vocabulary will exist solely in English, and (grand)parents will likely adapt their use of words to cater to the (grand)children. Wishing to help their (grand)children progress, (grand)parents seek to use and understand English words when speaking to them.[18]

In terms of the standard of English acquired, the learning of English in Pithoragarh is partial at best. Pithoragarh English is perpetuated by an education system in which the subject is taught in the lower years by semi-fluent speakers or non-speakers. By the time students reach the exam years 9th-12th, and are allocated fluent or semi-fluent speakers as subject teachers, they are facing textbooks several levels above their comprehension level, and many teachers resort to giving explanations in Hindi to ensure that students have a chance of scoring well in their exams. As a result, students leave school with limited fluency themselves, but plenty of English vocabulary. These are universal problems faced by schools in rural and semi-urban India, which have been well-documented by LaDousa (2014) and others.

What can we conclude about the type of influence(s) that may result from an area-wide increase in the proportion of children and young people learning English? The difference between learning English as a subject in an otherwise Hindi-medium educational programme and studying entirely in English is enormous, and taken at face value would result in a huge increase in borrowings entering the language of young people initially and then the general

---

[17] CBSE (Central Board for Secondary Education) is the most prominent English-medium school examination board in India and is managed by the government's Ministry of Education.

[18] The influence on the grandparent generation would be increased by the common local practice of living together in joint families, a residential situation in which sons and their wives and children continue to live at home with parents. At the time of recording, 8/47 research participants lived in a joint family, including two sets of sister-in-laws who recorded together.

population, as well as an increase in the propensity of the same children and young people to use English in their everyday language via a mode of speech characterised by CS. However, the situation in Pithoragarh is not altogether clear-cut, since in practice many English-medium schools in the town (even good ones) operate by using a mixture of Hindi and English in the classroom to make up for lack of English comprehension on the part of the students, similar to the style of teaching exemplified and analysed by LaDousa (2014). A pure English-medium education system relies on teachers being fully fluent in English, which most Pithoragarh-dwelling teachers are not, and therefore both teachers and students fall back on their language of greater proficiency in order to communicate effectively. This in turn influences both the kind and volume of English which students acquire and which they are then able to draw on as part of their linguistic repertoire when communicating with family and friends outside of school. This largely explains the difference between the heavily code-switched variety of Hinglish spoken by the upper middle class in Delhi and other urban centres and the much milder forms of code-switching employed by young people in Pithoragarh. In spite of this, the level of contact with English through education is unsurpassed by any other type of contact experienced by residents of Pithoragarh and is likely to have the biggest impact on Hindi-English mixing in future decades as the number of children attending English-medium schools increases and as the standard of English spoken by teachers also increases. It should also be noted that the standard of English in schools is incrementally improving and in half a generation the situation is expected to be quite different from what is described here.

3.3.5   Technology and the media

Technology and social media

In accordance with the growing trend across the country, there has been a huge increase in the number of people owning mobile phones in Pithoragarh between 2007 and the present day.[19] The late introduction of Hindi keyboards and Hindi language apps means that Facebook and WhatsApp are accessed and used in English by most users, despite Hindi now being available. In 2008, the only mobile phones available on the Indian market had English keyboards and Vaish claimed that 'only those with literacy in English can send SMSs' (2008:212). Contrary to this claim, in the experience of the researcher, literacy in 'Roman Hindi', with its much less restricted spelling conventions, is all that is required. The prevalence of WhatsApp and Facebook are responsible for increased literacy in Roman Hindi

---

[19] The researcher moved to live in Pithoragarh in 2007 and has personally witnessed the multiplication of mobile phone use in the town.

among the Hindi-medium educated population. Typing transliterated Hindi in Roman script is the norm for most app users, and there are a variety of spellings in operation. One example of this is the word for Yes, हाँ, which is variably spelt *ha, haa, han* or *haan* with the double vowel employed to represent the long vowel आ by some speakers (whereas others presume this is understood), and similarly the nasalization represented by the letter 'n' considered obsolete by some speakers but employed by others. The variety of spelling is not a reflection of any variation in pronunciation, but rather in the personal transliteration standards of each speaker, which are largely internally consistent, if idiosyncratic. Hindi keyboards became available much later and now applicants for certain local government jobs such as Clerk are required to sit typing exams on a Hindi keyboard, but this is not a common skill. In fact, until recently, students at Hindi medium schools were taught to use computers only with an English keyboard and to use Microsoft Word with English menus, despite the availability of Hindi software and even now, Hindi typing is not promoted.[20]

Film and TV

Bollywood films, TV serials and comedy shows are popular in Pithoragarh. Unlike western films, in which even a short utterance in a foreign language is translated with subtitles, utterances in languages other than Hindi (such as Punjabi, Bihari or another regional dialect) are accepted as part of the multilingual language mix and not subtitled. As argued in the previous chapter, despite an increasing proportion of code-switching with English, the content is still accessible. Cricket is the most popular sport in India, and widely followed by all members of society. Both sports channels and news channels are available in both Hindi and English. Despite rare Pithoragarh residents preferring to listen to cricket commentary in (Indian) English rather than Hindi, in the experience of the researcher Pithoragarh residents would generally choose Hindi channels and programmes over English ones. Hollywood films are not widely appreciated but are watched by some of the younger generation of English-educated students. It is with this broad perspective in mind that the questionnaire results presented below should be interpreted.

3.3.6   Use of English in Pithoragarh

Questionnaire results

English is incontrovertibly part of daily life and usage in some form for many of Pithoragarh's residents, although the extent of English usage differs widely. Research participants were

---

[20] Information received from research participant Bhagwan (2020, p.c.)

asked to select which activities they used English for from a list (see Appendix C for a description of the research participants and Appendix D for the sociolinguistic questionnaire). The number of participants who completed the questionnaire was 43, so each number is out of a possible total of 43. The results are shown in Fig 6. The tasks are divided into activities which requiring an active production of English and those which require a passive comprehension of English. Some of the tasks could have been interpreted as either requiring both comprehension and production or only comprehension of English, and these have been separated out into a third list 'Producing/Understanding English'.

Producing English

✓ Competitive exams	5% (2/43)
✓ Job interviews	2% (1/43)
✓ Singing	5% (2/43)
✓ Telling the kids off	2% (1/43)
✓ Writing	2% (1/43)

Understanding English

✓ Reading the newspaper	37% (16/43)
✓ **Using the computer**	**56% (24/43)***
✓ Watching films	40% (17/43)*
✓ Watching TV	44% (19/43)*

Producing/Understanding English

✓ Facebook	49% (21/43)
✓ In my job	26% (11/43)
✓ **Studying**	**53% (23/43)**
✓ Conversing with my family	26% (11/43)
✓ Conversing with my friends	47% (20/43)
✓ **Whatsapp/SMS**	**70% (30/43)**

Fig 6: Uses of English – results of questionnaire administered to 43 research participants

Results show that the activities which incontrovertibly involve producing English have an extremely low score (one or two participants). In terms of comprehension skills, more than half the participants (24/43) stated that they employed English for 'Using the computer'. As mentioned earlier in the section on technology, this is because software programmes have English menus and laptops and PCs are sold with Roman keyboards; it is not expected that half of the participants are carrying out word processing and other complex tasks in English. In addition to computer use, there were two other items marked with an asterisk (*): 19 participants who used English for 'Watching TV' and 17 for 'Watching films'. Contrary to

surface level interpretation, this does not entail that participants watch English language films or TV programmes. Rather, follow-on interviews with a number of participants confirmed that their response referred to the code-switching that has become commonplace (although still measured) in Bollywood and on TV shows.

The activity for which the highest number of participants reported using English was communicating via 'WhatsApp/SMS' (30/43 participants). Typing and reading a WhatsApp message generally requires knowledge of Hindi transliteration into Roman script (unless both parties use a Hindi keyboard, which would be very unusual). It may also involve comprehension of code-switches from English used by one's interlocutor. Like communicative activities, which may require passive or active knowledge of English, 'Studying' also requires the student to read English text and may require them to also write in English. Half the participants also claimed to use English for 'Studying'. Despite the ambiguity, the questionnaire data gives an indication of the ways in which English (and/or Roman script) features in the daily life of Pithoragarh residents.

Style of Hindi-English mixing

The mixed style of spoken language referred to as Hinglish in the previous chapter is not (yet) common in Pithoragarh. All Hindi is mixed, in that it contains a large number of English borrowings, but the frequent alternational code-switching described by Anderson-Finch (2011) and Bhatia (2011) is rarely attested. Young people who are studying (or have studied) at English-medium schools are the most likely to speak a style of Hindi with more frequent code-switches, but these tend to be insertional rather than alternational and the core of the syntax remains Hindi, which cannot be said for Hinglish. Regardless of the volume of code-switches, in the words of Annamalai (2001:172), the 'mixed language is considered by participants in the speech as a variety of their mother tongue and not as a variety of English'.

Lexical borrowings are common and are on a cline between more and less established, with some lexemes frequently used among younger people and others common to all speakers. A distinction can also be made between borrowings which have retained their 'original' semantic meanings from English and those whose meanings have changed. The rest of this thesis focuses on one highly established lexical borrowing, the English-origin lexeme *type,* which is used in colloquial Hindi with a variety of innovative meanings.

3.4 Conclusions

The remote geographical location of Pithoragarh and the relatively recent advent of English-medium schools are the two main reasons why English has not had such an extensive influence on spoken Hindi in the town as in the more highly developed urban centres of North India. However, the longevity of the contact and the increasing importance of English has resulted in established borrowings from English which are in common parlance, including lexeme *type*.

The following chapter (Chapter 4) provides an introduction to loanwords and to theories of lexical competition. These will form the background for our understanding of what has happened to *type* in Hindi.

4   Introduction to loanwords

4.1   Lexical borrowing: a long-term outcome of language contact

Language contact globally is known to result in different outcomes depending on the length and intensity of contact, and attitudes towards each of the languages and their perceived status relative to each other. Indian society is heterogeneous and different levels and types of language contact with English have resulted in different outcomes for different communities within a region and between regions.

One of the most common outcomes of intense and prolonged language contact is lexical borrowing, the adoption of lexemes, thence known as 'loanwords', from one language into another which become a permanent item in the lexicon of the so-called recipient language. Loanwords are typically morphophonologically adapted to the patterns of the recipient language, although they may retain phonemes which are also adopted into the language. Crucially, they are used by monolingual speakers, not only bilinguals. Loanwords are also termed "lexical borrowings" after the process itself, or "loans" for short.

Thomason's (2001) typological borrowing scale categorizes language contact environments based on sociolinguistic factors including the length and intensity of contact, the level of bilingualism of speakers in the community, and attitudes towards the other language(s). According to Thomason's (2001) typology, lexical borrowing is prevalent in all language contact contexts, from least to most intense. Lexical borrowing is usually understood to precede structural borrowing, or to mediate it, although exceptions have been noted (e.g. Munshi, 2010). The closest fit for India is the second category on Thomason's (2001:70) borrowing scale which is labelled 'Slightly more intense contact', a context which is characterized by the bilingual agents of change ("borrowers") being a minority among recipient-language speakers. Thomason's Category 2 holds true for most of Hindi-speaking North India. The types of borrowing to be expected, and which we can observe in Hindi throughout North India are: lexical borrowing of function words like conjunctions and "nonbasic" content words; in terms of structural borrowing, minor changes in the frequency of usage of current structures, but not alterations to the structures themselves. New phonemes are expected to be found only in loanwords and not influencing the phonology of the recipient language in general.

The Loanword Typology Project and corresponding publication by Haspelmath and Tadmor (2009) are testament to how common lexical borrowing is. English lexis itself consists of a

high percentage of borrowed words; of the 1000 most frequent lexical items in the British National Corpus, 32 are of Scandinavian etymology and 487 come from French and/or Latin, a result of the history of invasions from the Vikings and Normans respectively (Durkin, 2014:37). Stanlaw (2004) illustrates the key role played by English loanwords in Japanese culture and society, including the way they have supplied an alternative set of colour terms for all except primary colours.

Loanwords can be single words, compound nouns, or phrases of two or more words which become fossilized or conventionalized as fixed expressions in the recipient language. Nouns are regarded as the easiest lexical items to borrow on an implicational borrowability hierarchy (Matras, 2009).

4.2   English loanwords in Hindi

Colloquial Hindi is replete with borrowings from English, which are almost always phonetically integrated into Hindi. There are large numbers of phonemes which are pronounced (consistently) differently, notably 't' 'd' 'r' but also including the modification of vowel sounds to align to the Hindi phonological inventory (Masica, 1991). Inclusion in a dictionary is also not a reliable criterion for judging the degree of spread or acceptability of a loanword. Firstly, prescriptivist views on the language may prevent lexicographers from including words of foreign origin; secondly, dictionaries often lag behind spoken language and so words may be in common parlance long before they are captured by a dictionary. The Oxford Hindi-English Dictionary (OHED) also deliberately omits English loanwords whose form and meaning have not been substantially altered (aside from the phonetic integration described above), on the grounds that their meaning is easily discernible to the user (McGregor, 2018:x).

Examples of loans which have been borrowed from English into Hindi and are used universally include **nouns** such as 'phone', 'mummy', 'school', 'form' and 'tension'; **compound nouns** such as 'medical store' (=pharmacy), 'degree college' (=university), and 'heart attack', **verbs** such as 'start' and 'apply' and **phrases** such as 'you know' and 'miss you'. All English verbs that are borrowed are integrated into Hindi as the 'nominal' part of a light verb construction, typically followed either by transitive *karna* 'to do', e.g. *sign karna* 'to sign (sth)', or intransitive *hona* 'to be', e.g. *deposit hona* 'to be deposited'. Borowiak (2012) explains how light verbs containing English lexical items pattern slightly differently to their Hindi equivalent light verbs (for example, including a novel VERB + VERB combination).

Similarly, Edwards and Gardner-Chloros (2007) discover innovative light verb constructions found in the speech of Greek and Punjabi bilinguals.

4.3   Lexical borrowing and code-switching

Lexical borrowing is often contrasted with another outcome of intense contact between languages known as 'code-switching', the mixing of two or more languages within the same sentence or conversation by bilingual speakers.

Examples of code-switching in Hindi are shown in Figs 7 to 9 below. Fig 7 is an example of a single-word code switch, Fig 8 is an example of a mixed noun phrase and verb phrase with some Hindi and some English words, and Fig 9 is an example of a multi-word code-switch. Fig 9 is interesting as it is a calque on the English phrase 'to <u>give</u> someone freedom of choice'; the verb is also underlined in the example.

> *bilkul alag **culture** rahata hai na un logon ka?*
> 'Their **culture** is (lit. remains) totally different, isn't it?'
> **Fig 7: Mamta (F64), convo_b**

> ***life** ka har ek **moment enjoy** karna chaahiye*
> (We) should **enjoy** every **moment** of **life**
> **Fig 8: Akanksha (F17), convo_n**

> *mere gharwaale bahut jyaada **supportive** hain. unhone mujhe **freedom of choice** <u>diya</u> hai*
> 'My parents are totally supportive. They've <u>given</u> me **freedom of choice**.'
> **Fig 9: Bhagwan (M18), convo_e**

The relationship between lexical borrowings and (particularly) single-word code-switches has been a subject of considerable debate during the past 30 years. Gardner-Chloros (2009), Matras (2009) and Thomason (2001) describe code-switching and lexical borrowing as different ends of a continuum. On one end, loanwords are morphologically and phonologically integrated into the speech of monolinguals, and have replaced a native form. On the other hand, speakers intentionally exploit both (all) linguistic systems at their disposal by switching at will between one language or variety and another.

When a single lexical item is uttered regularly by monolingual speakers within a speech community and structurally integrated into the recipient language, it can be considered a borrowing, or an established loanword; such is the status of the noun *type*, the focus of this thesis.

4.4 Lexical borrowing and semantic shift

Semantic shift is a diachronic lexical process by which the (semantic) meaning of a lexeme changes over time, the synchronic expression of which is polysemy (Koch, 2016). Zalizniak (2018) and colleagues are responsible for over twenty years of work compiling the Database of Semantic Shifts (DatSemShift.2.0), which records semantic shifts in the world's languages.[21] Interesting examples include the following:

- **Shift #2804   gold → money**
  - o   Six realizations, of which two are a result of lexical borrowings
- **Shift #2185   to cut (tr.) → pain**
  - o   Nine realizations, of which one was derivational and eight were polysemic

The semantic shift from *gold* to *money* is significant for our purposes as two instances (which equates to two pairs of languages) recorded the change as a result of lexical borrowing. Indeed, it is often the case that loanwords develop a different meaning (or meanings) in the recipient language than is retained in the donor language. Considering the extent to which lexical borrowing occurs in the world's languages, and the focus on semantic shift within the historical linguistics field, it is surprising that there have not been more studies that focus on the relationship between the two. Nyota and Mareva's (2012) study on semantic change in Shona street lingo is one example of a recent study that focuses specifically on semantic change in loanwords. The authors use participant observation and questionnaires in order to ascertain how words have acquired new meanings when borrowed from standard Shona to the urban youth variety of Shona street lingo. One example discovered through the research is a shift in the meaning of loanword *makeke* 'cakes' which has come to mean 'anything nice… a thing, situation, positive attitude or breakthrough' (Nyota and Mareva, 2012:117).

The development of innovative meanings and uses of a loanword in its new lexicon can be immediate, '*during* the contact phase itself', as Koch (2016:56) states, giving the example of *sombrero*, a Spanish word simply meaning 'hat' which in English took on the narrowed meaning of 'Mexican hat'. However, it can also be gradual, as a result of the interaction between lexemes jostling to find their place in an overcrowded lexicon, to appropriate the competition metaphor, which will be introduced and discussed in the following sections.

---

[21] The database is located at http://datsemshift.ru/.

4.5   Lexical borrowing and functional specialization

When a loanword is integrated into the lexicon of a recipient language, it may fill a lexical gap, for example, *phone* in Hindi, or it may enter as a synonym of another lexeme (or lexemes), a distinction Onysko and Winter-Froemel describe as 'new concept' versus 'close synonym' (2011:1552). In this latter case, it enters into competition with native lexemes, which, according to the competition metaphor, is commonly understood to lead to two possible outcomes: **differentiation** (either strict or partially overlapping) and **substitution**. Durkin's (2014) analysis of the 100-list, a compilation of words most resistant to borrowing in the world's languages, reveals that in English, 12 of the 100 are borrowed words, and that these loanwords have had an effect on the meanings or uses of native lexemes, or both. Differentiation can occur in a variety of ways, such as by register or by context; as Durkin writes, 'Very often, an existing term shows semantic narrowing or specialization rather than complete loss, as for instance native *hide* following the borrowing of *skin'* (2014:406-7). Substitution can be complete, such as the replacement of English *nim* with Scandivanian-origin *take*.

Onysko and Winter-Froemel (2011) analyse the semantic-pragmatic meanings of English loanwords in the German news magazine Der Spiegel and identify semantic shift in several of those non-essential borrowings for which there is a close synonym in German. Whilst semantically equivalent to a German word, they are used in different ways or with a different scope of meaning (i.e. more or less specialized) than the English loan. An example of this is the formation of a verb *shoppen* 'to shop for leisure' from the English *shop*. German *shoppen* is differentiated from the hypernym *einkaufen*, which is used for daily shopping (2011:1558). Secondly, loanwords *Teenager* and *Kids* express different denotative and connotative meanings than their German equivalents *Jugenliche* and *Kinder. Teenager* often occurs alongside other anglicisms relating to 'technology, fashion, music and lifestyle' (2011:1562).

Oesch Serra's (1998) investigation into the use of French and Italian discourse markers by Italian migrants in French-speaking Switzerland reveals a unique patterning of usage whereby the three discourse markers *ma, mais* and *pero* have become functionally specialized, and are also combined in distinct ways to structure discourse.

4.6   Principle of attraction

De Smet et al. (2018) critically evaluate the metaphor of competition and focus on a third outcome previously unaccounted for by competition theory, that of **attraction** between lexemes. The principle of attraction explains the phenomenon whereby competing linguistic

items rather than specializing or replacing one another in fact become more similar, increasing the level of formal and functional overlap between them. The authors argue that pre-existing similarity is the cause of the attraction between lexemes: '[s]ituations of functional overlap invite analogy' (2018:217).

One example from English concerns the variation between the *-ing* form and the *to*-infinitive form of verbs following the verb *begin*. Whereas the distribution of *to*-infinitive uses historically preferred agentive contexts and *-ing* preferred non-agentive contexts, over time the two have become more functionally similar such that the proportion of both constructions being associated with non-agentive subjects has increased. This change is concurrent with the gradual substitution of *to*-infinitive uses by *-ing* uses, a process not yet complete.

De Smet et al's (2018) theory also accounts for differentiation and substitution. Differentiation is understood as a result of an expression's membership of a constructional family which exerts synchronic as well as diachronic influence. For example, the authors show how *begin* is part of a family of verbs which all prefer agentive subjects. Since 'constructions with *-ing*-clauses more generally seem to favour verbs associated with agentive subjects', it follows that *begin* is more likely to be followed by an *-ing* form of the verb than a *to*-infinitive (2018:218).

4.7   Conclusion

We have defined lexical borrowing as the permanent adoption of a lexeme (or phrase) from one language to another, distinguishing this (as far as is possible in very general terms) from the spontaneous appropriation of linguistic items from another language known as code-switching. There are several processes that impact and intersect with lexical borrowing, which must be understood in the context of the whole lexicon, based on an acknowledgement that synonymous lexemes interrelate ('compete'). The outcomes of competition between lexemes include differentiation (functional specialization), substitution and, notably, attraction, a process by which lexemes become more similar to each other over time.

In the following section the main "protagonist" of this thesis, the English loanword *type*, is introduced. The goal of this thesis is to begin to uncover what has happened to *type* since being borrowed into Hindi. Since *type*, as all lexemes, does not exist in isolation but interacts with other synonymous lexemes, this will involve an analysis of how it functions in relation to close synonyms.

Investigating the grammaticalization of four Spanish manner adverbs which have become alike over the years, Aaron (2016) asserts: 'Not only do forms compete with each other, but they also may **borrow the path another has taken**' (Abstract; **bold** text mine). This thesis considers the possibility that a force of attraction between equivalent words may trigger or accelerate a process of grammatical replication causing one word to become more similar to another and specifically that this may have occurred to loanword *type* in attraction to native Hindi lexeme *aisa*.

The journey of *type* in English (and as we shall see, in many other languages; see Chapter 7), itself a loanword, exemplifies not only the semantic shift discussed in this chapter, but a process of grammatical change known as grammaticalization. The following chapter provides an introduction to the grammaticalization of *type.*

5 Theoretical background to the grammaticalization of *type*

Type nouns, also known as taxonomic nouns or categorizer nouns, are a lexical device for categorising items within a (pseudo-)taxonomy or a hierarchy, e.g. 'A recliner is a type of chair'. The lexeme *type* designates a subcategory within a superordinate category, so in the example sentence given, 'recliner' is classified as a subcategory of the category 'chair'.

Type nouns cross-linguistically include lexeme *type*, from which the group of nouns earns its name, but also cognates and synonyms of *kind, sort, class, form, quality, (sub)species, style, and manner*.[22] Type nouns are highly polysemous and each of them has a (slightly) different semantic constitution. In English, the principal type nouns are *kind*, *sort* and *type*. Of these, *kind* and *sort* are more similar to each other and have developed many more additional functions than *type.* It is outside the scope of this thesis to consider why this is the case, or to review the development of all type nouns in each of the languages represented, therefore what follows will focus almost exclusively on *type* and its synonyms and cognates, with brief reference to other lexemes where relevant.

5.1 Early history of *type* in English

The taxonomic use of type nouns originated in European scholarly contexts, particularly influenced by Aristotelian logical philosophy relating to the essence and practice of categorization. For example, Aristotle's Book IV of The Topics deals with the definition of *genus*, including the difference between *genus* and *species*. Scientific taxonomic categorization was subsequently developed by scholarly institutions based on Aristotle's ideas, classifying and defining an object in relation to its taxonomic hierarchy consisting of a superordinate category and subordinate subcategories. From an origin in biological classification, type nouns then 'became part of the general lexicon early on, beginning with the most ancient texts in European vernaculars' Mihatsch (forthcoming a). Unlike the cognates of *kind, genus* and *class*, *type* does not feature in official scientific taxonomies. Rather, it emerged later as a categorizer for general everyday items. Of all type nouns, the taxonomic usage of 'type' is attested later than that of 'kind' and 'species' in Romance languages and in English.

The Historical Thesaurus of the Oxford English Dictionary (HTOED) traces the history of nominal *type* from the sixteenth century, when it first had the meaning 'symbol'. It

---

[22] List compiled from Mihatsch (2016) and Masini (2016).

subsequently developed a sense similar to 'pattern', 'model' (prototype) and 'exemplar'. The category meaning of *type* in English, equivalent to 'kind' and 'class', was first attested in **1854**. The HTOED quotes examples from 1854 and 1855 which show two different sentence structures (Fig 10); here and elsewhere in this thesis the *type* phrase is formatted in **bold** for emphasis and ease of viewing. The first *type* phrase 'a *type* of reason' follows the structure 'a type of N'. The second example shows the category noun *small pox* preceding a postponed *type* phrase, rendering the structure 'N of X type':

a) **1854** D. Brewster More Worlds iv. 73   On a planet more magnificent than ours, may there not be **a *type* of reason** of which the intellect of Newton is the lowest degree?
b) **1855** T. B. Macaulay Hist. Eng. IV. xx. 531   The Queen was sinking under **small pox of the most malignant *type***.

Fig 10[23]

In modern English, the primary usage of *type* is the categorizer sense, which has two subsenses. Current uses of *type* have been investigated in the Oxford English Dictionary online (OED) residing at lexico.com (Fig 11).[24]

---

type **n.**

1. A category of people or things having common characteristics.
    'this type of heather grows better in a drier habitat'
    1.1. *informal [with adjective or noun modifier]* A person of a specified character or nature.
        'two sporty types in tracksuits'
    1.2. (**one's type**) *informal* The sort of person one likes or finds attractive

---
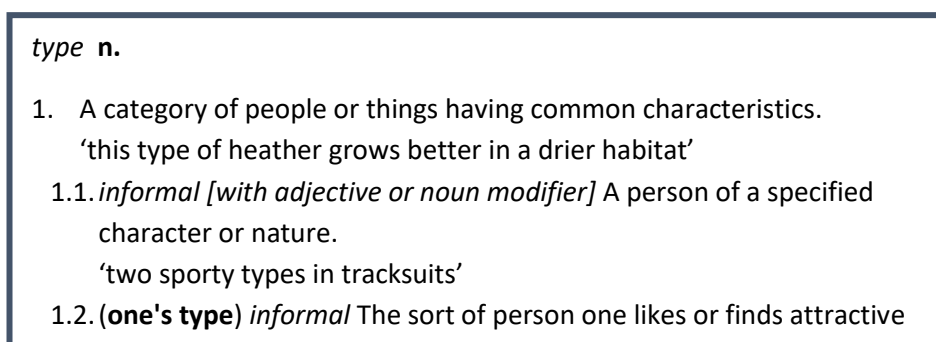
Fig 11: Lexical entry for *type* from OED online[25]

5.2   Grammaticalization theory

Grammaticalization is a process of language change whereby lexical items come to take on grammatical functions and grammatical items take on new grammatical functions. An example is the lexical verb 'to say' which has developed a use as a complementizer 'that' in West African language Ewe (Hopper and Traugott, 2003:13-15). However, such cases of grammaticalization where a lexical item develops into a grammatical item are not the most

---

[23] https://www.oed.com/view/Entry/208330? Accessed 22nd January 2020.
[24] Described as one of the "Oxford Living Dictionaries" on the website, this online dictionary was updated quarterly with new lexical entries, sub-entries and senses. Sadly, the online version of the dictionary is being discontinued from 26th August 2022.
[25] https://www.lexico.com/definition/type. Accessed 17th January 2022.

commonly-attested; more frequently, a linguistic item already performing a grammatical role develops a further or different role, which may also involve a change of grammatical category (Heine and Kuteva, 2005). An example of this is the tendency for indefinite articles to be formed from grammaticalized cardinal numeral 'one', a pattern attested in numerous languages.

Grammaticalization theory was propounded by Lehmann (1982, 2015) and developed by Hopper (1991), Hopper and Traugott (2003), Heine and Kuteva (2005) and others. Heine and Kuteva (2005:15) list four interrelated mechanisms of grammaticalization (which they term 'parameters'), reproduced here as Fig 12:

a) **extension**, i.e. the rise of novel grammatical meanings when linguistic expressions are extended to new contexts (context-induced reinterpretation)
b) **desemanticization** (or "semantic bleaching"), i.e. loss (or generalization) in meaning content
c) **decategorialization**, i.e. loss in morphosyntactic properties characteristic of lexical or other less grammaticalized forms, and
d) **erosion** (or "phonetic reduction"), i.e. loss in phonetic substance

Fig 12: Reproduced from Heine and Kuteva (2005:15; bold characters mine)

The **extension** of meaning of a lexical item arises through semantic-pragmatic reinterpretation as a linguistic item is applied in a new (pragmatic) context. The process of change depends upon a **bridging context**, a linguistic construction which can be analysed syntactically in two different ways. This allows the structure to be interpreted differently and reapplied in a new pragmatic or syntactic context (or both) and thereby acts as a pivotal point for a linguistic item changing its meaning and its grammatical status. The resulting grammatical innovation (if it catches on) is subsequently propagated throughout the speech community. Hopper and Traugott claim that 'reanalysis and analogy are the major mechanisms in language change', with reanalysis related to the initial innovation and analogy related to its spread (2003:69).

Mechanisms (b-d) involve both loss and gain. Grammaticalization would ideally involve all four mechanisms/parameters, but it does not always; in certain cases, it involves only **desemanticization** (Heine and Kuteva, 2005). The process of desemanticization occurs as the lexical meaning of the linguistic item is reduced ('bleached') and it takes on a grammatical function. The new use is more widely distributed than the previous one had been, since the reduction in lexical content makes the linguistic item applicable to more contexts.

Some scholars have argued that desemanticization is somewhat a misleading term as the use of a linguistic item in new contexts gives rise to new semantic meanings. In fact, Hopper and Traugott (2003:98) confirm that '[t]he process of **demotion of some lexical meanings and promotion of others** is characteristic of semantic change in general' (**bold** mine). Lexical meaning is complex and nuanced. During the process of grammaticalization certain aspects of the meaning of a lexical item are foregrounded and others backgrounded. Promoted meanings tend to be abstract (Heine and Kuteva, 2002) and can veer towards subjectivity. As these foregrounded meanings are applied to more pragmatic contexts (via extension), the new nuanced meaning becomes conventionalized. In particular, scholars have found that throughout the process of grammaticalization (and in fact common to semantic change in general; Traugott and Dasher, 2001:89-90), the semantic meaning of a linguistic item increases in subjectivity and expression of the inner cognitive thoughts or attitudes of the speaker, a concomitant process known as **subjectification**.

5.2.1   Grammaticalization involves semantic and syntactic change

It will be evident from the above that **semantic shift**, which we defined earlier as a diachronic process of lexical semantic change, plays a key role in the grammaticalization process. Although semantic shift can occur without grammatical change, the reverse is not true: semantic shift is a requisite component of grammaticalization (Hopper and Traugott, 2003).

A brief mention must be made of the distinction between semantics and pragmatics, which can be roughly defined as the difference between the meaning of a word versus its meaning in interaction. Despite the complexities of both disciplines, there is a high degree of interconnectedness between the two. Portner (2005) researches the connections between semantics and pragmatics as well as their relationship with syntax and concludes: 'it turns out that semantics and pragmatics are interwoven in a way that makes it impossible to really study semantics without simultaneously doing some pragmatics, and vice versa.' (Portner, 2005:176) It should be apparent that both are intimately related to the process of grammatical change, particularly to the extended meanings of type nouns, some of which are highly pragmatic in nature. In this thesis the two will be largely considered together, such that the grammaticalization parameter of **extension** which refers to the linguistic item being employed in a new context will be understood to mean a new semantic-pragmatic context.

A further element of grammaticalization is the evolution of linguistic items from one grammatical category into another, a process known as **decategorialization**. The term focuses on the loss of morphosyntactic properties associated with category membership of

'more lexical' categories such as the noun, although of course the process simultaneously involves the gain of characteristics associated with new grammatical functions (Brems, 2011:113). In Hindi, the use of lexical verbs to form the second verb in a verbal compound is an example of decategorialization. Hopper and Traugott (2003:112-114 following Hook, 1991) place this development on the grammaticalization pathway between full lexical verb and auxiliary – Hindi light verbs such as *lena* 'take' and *dena* 'give' can function as both main verbs and light verbs. As light verbs, they still carry some semantic content, particularly perfectivity, but no longer function as the main content verb in the clause.

The fourth mechanism of grammaticalization is the process of **erosion** during which the linguistic item is phonetically reduced. This is particularly common at word or morpheme boundaries (Hopper and Traugott, 2003:157). A pertinent example is the loss of the sound /v/ and the change in vowel sound from /ɒ/ to /ə/ in *of* as the two words *kind of* are fused and become *kinda*. Erosion does not always occur; Schiering (2010) investigates the outcome of cliticization during the process of grammaticalization cross-linguistically and argues that erosion is a normal and expected outcome of grammaticalization for stress-based languages, but not for syllable-based or tonal languages.

5.2.2   Unidirectionality and grammaticalization pathways

A central tenet of grammaticalization theory is that the process is **unidirectional** - linguistic items progress along a cline from less grammatical (more lexical) to more grammatical and not the other way around. This has been shown to be the case regardless of whether the origin of the change process is language internal or contact-induced (Heine and Kuteva, 2005). Grammaticalization progresses along a step-by-step fashion, meaning that '[s]ynchronically, this fact can be described in the form of an implicational scale: If a given article has stage X then it also has all preceding stages.' (Heine and Kuteva, 2010:93)

Scholars have identified cross-linguistic regularity in the way that certain lexical or grammatical items develop into other grammatical items. These cross-linguistic patterns have been termed "pathways" (Hopper and Traugott, 2003), and range from more general to highly specific progressions. One such grammaticalization pathway is presented in Fig 13.

Noun (body part) > Noun (spatial relationship) > Adv > P > Case affix

Fig 13: From Hopper and Traugott (2003:6)

Heine and Kuteva and Kuteva et al's (2002, 2019) 'World Lexicon of Grammaticalization' first and second editions contain an encyclopaedic list of cross-linguistically attested

grammaticalization pathways. These pathways have all been identified in more than one language and preferably more than one language family.

5.2.3 Layering and persistence

The concepts of 'layering' and 'persistence', both coined by Hopper (1991) are also essential for our understanding of grammaticalization. The concepts are interrelated: 'layering' refers to the enduring presence of older (less grammaticalized) uses and forms of a linguistic item alongside newer ones. This manifests itself synchronically as polysemy. 'Persistence' is also concerned with longevity but refers to the tendency for newer forms of a linguistic item to retain part of the lexical semantic meaning of an earlier form. This can manifest itself synchronically by restricting the uses to which a linguistic item is put. Hopper and Traugott (2003:96) share how 'details of its lexical history may be reflected in constraints on its grammatical distribution', giving the example of the difference between future forms *will* and *going to* in English, which is attributable to their origins as 'directional' and 'progressive' lexemes respectively.

5.2.4 Discussion on the use of synchronic versus diachronic data

Grammaticalization has been studied from both a synchronic and a diachronic perspective. Synchronic data, such as that analysed in this thesis, represents a snapshot in time and shows us how lexemes and constructions are being used, by whom and in which ways, at that specific point. However, it does not on its own provide evidence for what the linguistic item was in its incipience or how exactly it developed. Answers to these questions are supplied by diachronic data, the study of which has enabled scholars to track subtle changes in the syntax and semantics of linguistic items over time and identify pathways of grammaticalization.

A number of scholars such as Rosenkvist and Skärlund (2013) have questioned the reliability of claims about grammaticalization made on the basis of synchronic data alone. Their arguments centre on the fact that co-existence does not equate causation, and that comparable diachronic data is the only solid evidence allowing us to 'state with certainty that a change has taken place (A > B)' (2013:314). Studies that include such rigour include the authors' own detailed analysis of the grammaticalization of lexeme *typ* in Swedish, and a diachronic study of non-nominal *tipo* in Italian by Voghera (2013).

However, other scholars such as Heine and Kuteva (2002), whilst acknowledging that grammaticalization is a historical process and can only be fully confirmed with access to historical data, maintain that if such data is unavailable then known cross-linguistic grammaticalization processes can be referenced to help reconstruct a grammaticalization

process. Moreover, the nature of layering in grammaticalization is such that older linguistic forms and meanings often coexist with newer ones, so synchronic data can offer us a window into the past.

Both synchronic and diachronic data are valid and necessary. A comparison of diachronic studies on similar linguistic items cross-linguistically allows us to identify cross-linguistic patterns, including common semantic shifts and grammaticalization paths. The detailed diachronic analysis carried out by scholars on languages for which substantial data is available can be used as a reference for hypothesizing and reconstructing grammaticalization processes in languages for which diachronic data is unavailable. This is the approach taken in this thesis.

5.2.5   Grammaticalization as distinct from polysemy copying

It was established earlier that grammaticalization depends upon semantic change, specifically the emergence of polysemy. However, grammaticalization is distinct from **polysemy copying**, the phenomenon whereby a lexeme from one language which has one sense in common with a polysemous lexeme from another language takes on one or more additional senses from the polysemous lexeme as a result of language contact.

Heine (2012) argues that grammatical functions cannot be copied in the same way as lexical meanings and that what appears to be polysemy copying is in fact the replication of a use pattern followed by the normal grammaticalization process. The existence of a polysemous grammatical morpheme in one language which is in close contact with another triggers speakers to start using the morpheme in an innovative way, and this leads to associated grammatical changes in line with the usual process of grammaticalization.

5.2.6   Grammaticalization and pragmaticalization

The recent focus of attention on the development of discourse markers has led some researchers to highlight apparent incompatibilities between the progression of discourse markers towards an increasingly discourse-based function independent of the grammar of a sentence, and the main tenets of grammaticalization theory. Heine (2018:42) summarises the elements of discourse marker development which are at odds with grammaticalization theory as follows: 'movement outside the syntax of a sentence, from prosodically integrated to non-integrated status, and from meanings as part of the sentence to metatextual meanings or functions'. These theoretical conflicts have caused some scholars to question whether grammaticalization is the right framework for understanding the process of change. Ocampo (2006) presents an alternative explanation he terms **discoursivization**. Others like

Diewald (2011) have preferred to use the term **pragmaticalization** to describe a specific kind of grammaticalization relating to the development of pragmatic functions. Aijmer (2002) uses both terms throughout her volume, but ultimately resorts to the more commonly used term grammaticalization to explain the development of discourse particles throughout.

One problem with the proposal of a new theory or new terms to explain the phenomenon is that the first stages of the development of discourse markers follow grammaticalization principles and processes exactly. The conception of an entirely new process seems to be theoretically redundant when the change from Categorizer to Similative, for example, is a clear case of grammaticalization. Rejecting the idea that the development of discourse markers is a distinct process from grammaticalization, Heine (2018) argues instead that discourse markers are **grammaticalized theticals**. He proposes that the progression from lexical item to discourse marker involves two periods of grammaticalization, between which linguistic items are **coopted** for use as discourse theticals. Heine (2018:40) defines the process of **cooptation** as 'an instantaneous transfer of clausal, phrasal, or other pieces of Sentence Grammar to Thetical Grammar for metatextual functions'.

> (Grammaticalization) > Cooptation > Grammaticalization
> **Fig 14: From Heine (2018:41)**

Following Heine (2018), the term **grammaticalization** will be used throughout this thesis to refer to the end-to-end process (Fig 14), with the understanding that cooptation may have occurred.

5.2.7   Conclusion: Grammaticalization triggers

The theory of grammaticalization has been thus far presented as a process of language change whereby linguistic items lose their lexical meaning and syntactic features in stages as they gain grammatical meaning and the syntactic features associated with their new function. This process of change has been identified in scores of languages and explains dozens of phenomena. At a micro level, the process is enabled by bridging contexts which facilitate, or even provoke, semantic reinterpretation and syntactic reanalysis. However, this does not explain the macro level question of why grammaticalization occurs and what triggers it. All language change is a result of speaker innovation. The primary trigger for grammaticalization is often understood to be universal cognitive processes such as metaphor and metonymy (see Hopper and Traugott, 2003; Koch, 2016). Whilst initially developed as a theory explaining language-internal change, it has since been recognized that grammaticalization can also be triggered by an external influence such as language contact.

Additionally, the complimentary theory of **attraction** presented in Chapter 4.6 offers insight into the relationships between synonymous or partially synonymous lexemes within a language and may also prove a trigger for grammaticalization.

Heine and Kuteva's (2005) landmark work takes as a starting premise that language contact is one possible cause of grammatical change alongside language-internal grammaticalization. Grammatical change as a result of contact, also termed grammatical replication, involves the replication of grammatical use patterns possibly also resulting in the innovation of replica grammatical categories in the recipient language on the basis of the model language. Moreover, the authors argue that contact-induced grammaticalization fulfils exactly the same parameters of "ordinary" grammaticalization, in other words, the nature of the grammatical changes which occur as a result of contact are no different than those which occur without any external influence. This begs the question as to whether the changes would have occurred anyway had contact not occurred. The authors address this question in their volume, responding that grammaticalization can be (partially) attributed to contact if the influence of contact with another language seems better able to explain the change, or the speed of change. A similar approach will be followed here.

5.3   Role of the speaker and hearer in borrowing and grammaticalization

5.3.1   Individual speaker agency – speakers and innovators

Ultimately, the process by which words spread and then become integrated into the (active) vocabulary of a population must be attributed to the agency of the individual speaker. We established earlier that attitudes towards English in India are almost universally positive, primarily through necessity, but also as a means of raising one's status in society. The motivation to choose words of English origin is therefore a corollary.

The journey of a loanword (such as *type* in Hindi) follows several stages, which we can divide into the <u>borrowing process</u> and the <u>grammaticalization process</u>. The borrowing process consists of the initial "borrowing" of the lexeme and its subsequent propagation or spread in the community. The grammaticalization process entails use of the lexeme in a semantically ambiguous context (the bridging context), leading to its semantic reinterpretation and usage in another syntactic environment. This change is then also propagated throughout the community.

5.3.2   The role of the speaker

Throughout the entire process, there are multiple points where speaker agency is pivotal. There are also multiple factors influencing speaker agency, including: level of bilingualism; exposure to patterns of use in the model language, the influence of which will be mediated by the level of bilingualism; and speaker attitudes towards the language, which can lead to intentionality in the use of the language.

Matras (2009) argues that some level of bilingual competence, however limited, is required in order to use a word from another language (following a broad definition of "bilingual"). The meaning assigned to a word is more likely to be closer to its original meaning (in the donor language) the higher the speaker's proficiency. The speaker who appropriates a word from another language is the "innovator". The word may be used in conversation by a speaker only once and never again, or it may be a word the speaker uses regularly, in essence, an established loanword in their idiolect. In either case, for the lexeme to become established as part of the lexicon of the language, it must be propagated throughout the speech community. There is no way of capturing the first instance of a lexeme or phrase being used in discourse; with a large volume of written texts, it is possible to identify the first attestation date to some degree of certainty, but with spoken data, it is almost impossible. Even if the first instance were recorded, by some amazing fluke, there would be no way of knowing that it was indeed the very first usage. As Thomason writes, '[I]t is important to

recognize that our chances of tracking the full course of any linguistic change, from the first speaker's initial innovation to the spread of the innovation throughout a speech community, range from slim to none.' (2001:130)

5.3.3 The role of the hearer

The agency of the speaker is critical in propagating the usage of a borrowed lexeme, but crucial to the process is an intermediate step whereby an interlocutor in a conversation (the "hearer") hears the lexeme or phrase in use by the innovator and interprets its meaning. The hearer then uses the word in a similar semantic-pragmatic context with the meaning which he or she had understood (Koch, 2016). This process of semantic interpretation may give rise to subsequent innovations if the hearer misinterprets the intended meaning of the speaker and promulgates a slightly different usage of the lexeme unintentionally. These further innovations are likely to be subtle, since by and large speakers having learnt a new word intend to use it with its "correct" meaning as they understand it. When enough speakers use a word and it is heard, interpreted and repeated by hearers who later function as speakers, the word spreads to monolingual speakers with no exposure to the donor language. Backus (2005) theorizes dissemination of linguistic forms using the concept of intentionality. Once a novel lexeme has spread to a monolingual speaker and they start using it, we can say that the proliferation of the language change is unintentional. Monolinguals' appropriation of the word is a key part of the process, although they may remain unaware that the word originated from another language.

5.3.4 The role of the speaker and hearer in grammaticalization

An established loanword is part of the lexicon of a language and is therefore eligible for participation in change processes that ordinarily affect words and phrases in a language, including semantic change and grammaticalization (which includes semantic change, as highlighted above). The beginning of any process of grammaticalization is the use of a lexeme (or phrase) in a semantically ambiguous context (Step (a) in the diagram below). This prompts semantic reinterpretation on the part of the hearer. Koch (2016) illustrates the two possibilities wherein either the speaker or the hearer is the primary agent of change. These possibilities are described incorporating Backus's (2005) concept of intentionality.

In Step (a), it is possible that the speaker acts intentionally to use a lexeme in a novel way, exploiting the syntactic ambiguity that allows multiple semantic interpretations. It is also possible that the speaker unintentionally creates ambiguity, allowing the hearer to cognitively (re)interpret a lexeme (Step (b)). This semantic reinterpretation is coupled with

syntactic reanalysis, so that the morpheme, lexeme or construction is 'parsed' in a novel way. Steps and (a) and (b) thus constitute the bridging context discussed earlier. When the hearer becomes a speaker, he or she uses the lexeme with its innovative semantic meaning in the new syntactic context (Step (c)). As others hear and propagate this innovative usage, it becomes normalized and eventually established in the language. Fig 15 summarizes this process in a step-by-step format with a focus on the role of the speaker and hearer.

<u>Grammaticalization process</u>

a) use of the lexeme in a semantically ambiguous context --> any speaker or innovator
b) semantic reinterpretation and syntactic reanalysis --> any speaker/hearer
c) use of the lexeme with its new meaning and syntax --> any speaker
d) propagation of the new (form and) meaning throughout the community --> all speakers
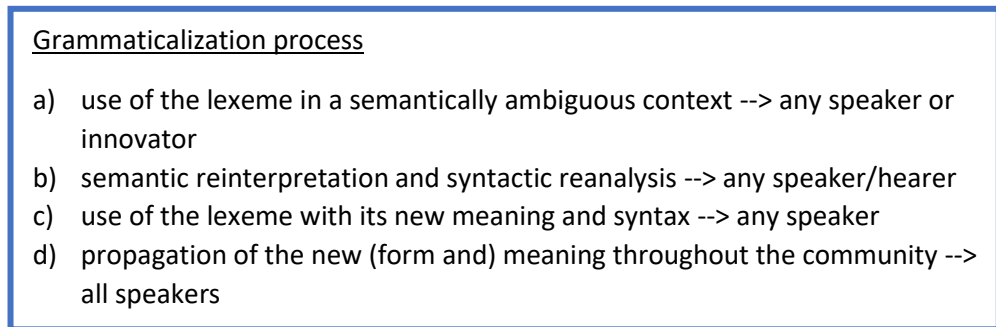
Fig 15: Step-by-step description of the grammaticalization of a loanword

Having presented the expected process and outcomes of lexical competition and language-internal grammaticalization at a high level, the following chapter presents the methodology used to investigate the functions of *type* in the colloquial Hindi of Pithoragarh residents.

6   Methodology

The chapter begins with an explanation of the researcher's years in Pithoragarh from the perspective of role, community involvement and language use. This background inevitably underpins the rest of the thesis and contributes to various aspects of the methodology. The primary data on which this thesis is based is a collection of recorded spoken conversations from Pithoragarh. The processes utilized for recording, transcribing, extracting and analysing the data are covered in detail in this chapter. This includes the analysis of innovative uses of English loanword *type* as well as the analysis of equivalent lexemes in Hindi which can be considered (partial) synonyms. A sentence acceptability test was used to validate that the recorded data was within the normal range of spoken language for the community. An additional data set consisting of informal web-based data was also analysed with a view to corroborating findings. In all cases, analysing the context of the utterance containing the target lexeme was key to understanding its meaning.

6.1   Researcher's role in the community

The researcher lived and worked in Pithoragarh full-time from 2007 to 2013 and then for part of each year until 2016. The researcher participated in all aspects of community life from festivals to family events (births, deaths and marriages) and became fluent in the local dialect of Hindi (spoken and written). As a job, the researcher taught English at the language school ILC and at various local schools on a contract basis through the ILC. Over the nine years, she was also involved in management, staff training and business development. Between 2014 and 2016, the researcher had a dual role as both a researcher and a teacher, conducting field research for her Master's thesis and this doctoral thesis alongside teaching responsibilities.

The researcher's modus operandi was to speak Hindi with everyone in the town with the obvious exception of colleagues and students at the language school with whom she spoke both English and Hindi. The researcher's language mixing style was on the conservative side, utilising more Hindi and fewer English words than the more innovative local speakers. She speaks and understands some Kumaoni.

Since leaving Pithoragarh in 2016, the researcher has maintained relationships with many members of the community and has regular (daily or weekly) contact with 30+ people through WhatsApp and phone conversations. Although an ethnographic methodology has not been followed per se, the researcher's extensive and ongoing interaction in Hindi and her experience at the English language institute inform this study in an ethnographic way,

providing insight into Hindi language usage and language mixing as well as personal experience of English use and proficiency in the town.

6.2   Primary recorded data

The primary data on which this thesis is based is a set of audio recordings made between May 2015 and June 2016 in Pithoragarh. One of the aims of this thesis was to capture current forms of language use and take this naturally occurring data as the basis for analysis. Recordings were all made with a Zoom H6 digital recorder on a tripod, usually placed on a coffee table equidistant from the speakers.

6.2.1   Research participants

Research participants were selected via the friend-of-a-friend sampling method (Milroy and Gordon, 2008). An individual known to the researcher (the "contact participant") was asked to participate, and they chose friends, neighbours, or relatives that they would usually converse with and invited them to join in. The relationship between the contact participant and the researcher varied; two young men were colleagues and the rest were neighbours, friends or acquaintances. Some of the participants recruited via the friend-of-a-friend sampling method were unknown to the researcher before the recordings. By ensuring that speakers were recorded alongside people they regularly conversed with the likelihood of the recorded data reflecting their normal speech patterns was vastly increased. Participants all signed an ethical agreement form on the day of the recording, except in a few cases where neighbours or relatives entered the recording 'scene' and joined in the conversation uninvited.[26] In two cases, the researcher was unaware of their presence until much later (during the data transcription phase) and permission to use their recorded speech was therefore obtained retrospectively. The permission form is shown in Appendix A.

Participants are all residents of Pithoragarh (N=47). There are 16 male and 31 female participants reflecting different age groups and different levels of English proficiency. The number of participants from each age category is as follows (age at time of recording): **15-24** N=12, **25-44** N=21, **45-64** N=13, **65-84** N=1. The mother tongue or L1 of participants is either Hindi or Kumaoni, or in three cases, Punjabi (although these participants had been living in the area more than 30 years and were also fluent in Hindi).

---

[26] Pithoragarh culture is highly social and people leave their doors open for neighbours to walk in. It is also highly inclusive, so research participants would have naturally included guests in their conversations.

The speakers represent various elements of Pithoragarh society both in terms of their demographics and experience with English. Participant occupations included students, housewives, teachers, shopkeepers, a local businessman and a government employee. A sociolinguistic questionnaire was used to gather data on the participants' personal details, linguistic background and language use at the time of recording. The questionnaire was developed in Hindi using Devanagari script, which all participants would have been comfortable reading, and is reproduced in Appendix D. The data gathered contributed to building a holistic view of the participants.

6.2.2   Recording process

The contact participant for each recording group was asked where they would like to record. In the majority of cases, they chose to record at the home of one of the participants. Most recording sessions lasted an hour, although the recording may have consisted of more than one audio file recorded consecutively. On a couple of occasions, the participants were very busy, and recordings were split over two days. One of the participants from convo_j was a shopkeeper at a busy sweet shop. He and his friend asked if they could record whilst out for their Sunday drive, which was the environment in which they usually conversed. The recordings were cleaned up to remove the noise of traffic and wind using Audacity(R) recording and editing software.[27]

When giving instructions for the purposes of the data collection, the researcher was careful not to use English-origin words except where unavoidable, e.g. use of the light verb *record karna* for the verb 'record'. This was to avoid any priming effect or accommodation to the researcher on the part of the participants.

The title of the research project was given to participants on the permission form as पिथौरागढ़ के रहने वालों की हिन्दी 'Pithoragarh residents' Hindi'. Participants were all told the researcher was interested in how people speak everyday Hindi and asked to converse as they normally would with the people in their recording group. None of the participants were informed about the researcher's interest in lexical borrowing and code-switching in order to ensure that their language use would not be swayed towards a greater use of English-origin lexemes than normal. In a couple of cases, participants overcompensated by actively trying to avoid English-origin lexemes and speak 'purer' Hindi. In both convo_e and convo_f, there is an instance near the beginning of the recording where one speaker uses English words and

---

[27] Audacity® software is copyright © 1999-2020 Audacity Team. The name Audacity® is a registered trademark of Dominic Mazzoni. The version used was version 2.2.2.

another speaker whispers a reminder to use Hindi. This deliberate monitoring may have resulted in fewer code-switches to English,[28] and possibly fewer loanwords too (depending on the speaker's level of awareness as to the origin of the loanword). This suggests that the lexemes which 'slipped through' the filter are either so well established as loanwords or (in the case of convo_f) that the participants became so relaxed that they forgot to filter out English-origin lexemes. Other recordings show no signs of monitoring and the recorded language use of those participants known to the researcher is comparable to their normal language use. Since the researcher was interested in naturally occurring English-origin lexemes, it was imperative that this not be contrived; an end result of (possibly) slightly fewer English words due to the self-monitoring of a small number of participants is a much better outcome than the alternative, an overinflated use of English words.

Another outcome of the request that participants use their everyday Hindi was that one group of three neighbours switched to Kumaoni, their L1, soon after beginning the recording. Interestingly, the recording contains **31** English-origin tokens, all loanwords, including one token of *type.*

There was one recording which was removed from the final corpus after consideration. Recording convo_c was recorded at a place of work and was a conversation between two *aayas* in their office.[29] The recording was arranged by an acquaintance and carried out without the researcher present. However, when the researcher returned to the location to complete the post-recording questionnaire, one of the participants did not seem at all comfortable with the scenario. It became clear that she had signed the permission form at the request of her superiors and not due to a genuine desire to be involved in the project. This compromised both the main premise of the thesis, the capture of naturalistic data, and the ethics of the project, which demanded that speakers participate freely and of their own volition. For these important reasons, the recording was withdrawn from the project and the decision was taken to ensure all future recordings took place outside of the workplace to avoid unwittingly placing pressure on participants.

The outcome of the recording process is a collection of 23 recordings totalling 17 hours of recorded conversation from 47 participants. The recordings when grouped together can be

---

[28] In fact, this was confirmed by Bhagwan, an old colleague of the researcher and the contact participant from convo_e.
[29] The job of an *aaya* involves aspects of cleaning, making and serving tea, fetching files and assisting office clerks with other menial tasks.

considered a corpus according to its broad definition as a 'collection' of recordings under similar conditions and with residence in Pithoragarh as a common denominator.

6.2.3 Transcription

Recordings were transcribed into transliterated Hindi in the Roman script and using the software programme ELAN.[30] Since Romanized Hindi is a common means of written communication through mediums such as SMS, WhatsApp and Facebook, this approach is in line with modern usage (see Chapters 2 and 3). As there is no single standard for transliterated Hindi spellings, but rather a set of options based on certain principles related to vowel length and historical transliteration protocols, spelling conventions were determined by the researcher. Effort was made to ensure consistency of spellings both within and across transcriptions so that ELAN searches could retrieve all relevant words and phrases. The first pass of transcription was completed by participant Anita (convo_a and convo_b), who acted as a research assistant during the first phrase of the project; in a few cases, initial transcription was completed by another friend. All transcriptions were then checked by the researcher and updated where necessary to correct inconsistencies or inaccuracies.

Areas that are inaudible have been marked as xx. Reasons for inaudibility include speaker overlap; too soft or too rapid speech; and traffic or other extraneous noise. There were a total of 428 tokens of xx in the recordings. Where the inaudible speech was in close proximity to the use of a key lexeme, additional effort was taken to interpret the text, including creating short clips of the recording using the software programme Audacity and sending them to a trusted advisor via WhatsApp or email. In most cases, they were sent to a research participant, and wherever possible, one of the participants of that very recording.

6.2.4 Extraction of *type* tokens

A Word List generated from within the ELAN software permits an analysis of the lexemes used in the corpus of recordings. The number of English-origin tokens in the corpus is 11,001 out of a total token count of 183,084. This means that English-origin words amount to 6% of all words uttered. The Top 20 English-origin lexemes in the corpus are printed in Fig 16, along with their overall position in the list of all lexemes in the corpus. All of the lexemes in the Top 20 list are established borrowings in Pithoragarh Hindi with the exception of #17 *life*, whose 63 tokens came overwhelmingly from two recordings (convo_m and convo_n) and from

---

[30] ELAN (Version 5.3) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

speakers under the age of 35.[31] The eleventh most frequent English-origin word in the corpus was the noun *type*. This demonstrates its importance in the local variety of Hindi, which is one reason for its selection; another is the opportunity to contribute to recent research on innovative uses of taxonomic nouns cross-linguistically, as has already been highlighted in the previous chapter.

EOL position	Lexeme	# Tokens	Overall position
1	mummy	200	142
2	time	197	144
3	phone	166	168
4	school	136	205
5	papa	128	218
6	aunty	90	271
7	number	90	277
8	army	89	275
9	suit	79	307
10	photo	75	316
11	**type**	**72**	**322**
12	result	71	330
13	twelfth	71	331
14	teacher	68	343
15	paper	66	348
16	pass	66	353
17	life	63	359
18	minute	62	363
19	tuition	57	385
20	first	56	386

Fig 16

All utterances containing the lexeme *type* were identified with a search in ELAN, exported into a CSV file and opened in Excel for further analysis.

There were 72 tokens of the word *type* in the corpus, of which 5 were of *type* being used as a verb, forming part of a light verb construction with the verb करना *karna*, as in Fig 17. These 5 tokens were removed from the sample, since the subject of study for this thesis is <u>nominal</u> *type* and its extended uses.

hin	अभी	**टाइप**	कर	रहे	होंगे	रोल	नंबर
tra	abhi	**type**	kar	rahe	honge	roll	number
gls	right now	**type**	do	PROG	must.PL	roll number	

---

[31] It seems that both from the data and in the experience of the researcher, English-origin noun *life* is in the process of becoming an established loanword but has not reached that status yet.

eng   [He] must be **typing** the roll number right now

Fig 17: Lokmanyu (M19), convo_h-3

The remaining tokens numbered 67, which gives a normalized frequency of 0.37 (tokens per 1000 words) for this lexeme in the corpus. This is the primary set of utterances which were analysed thoroughly in context, as explained in the following section. There were no tokens of English-origin *kind* or *sort* in the corpus. The transitive verb *sort sth out* has been adopted into the speech of the younger generation and occurs fairly frequently,[32] but was not attested in the recordings.

6.2.5   Limitations of recorded data set

Allowing participants to choose their own conversation partners and record in a comfortable and familiar environment has resulted in highly naturalistic data. Although *type* was one of the most frequently spoken English-origin words in the Pithoragarh corpus, the number of tokens was still fairly small (N=67). The diachronic corpora C-ORAL-ROM (2005) analysed by Mihatsch (2016) consists of around 300,000 words in each of the four Romance languages (FR, IT, PT, SP). The total number of tokens of *type* cognates ranges from N=40 for French to N=262 for Italian (Mihatsch, 2016:139, Table 3). 67 tokens is thus comparable with the number of tokens used to analyse *type* in French. However, it is lower than the number of tokens in the other languages studied.

Since the 67 tokens would be divided into different categories of usage following analysis, and possibly still further into different structural patterns, there would not be the luxury of many examples upon which to build a hypothesis. In order to augment the data set and verify the hypotheses generated by the analysis of the synchronic data from Pithoragarh, a number of approaches were taken. One was to conduct data validation via an acceptability test to verify that the language contained in the original recordings is typical of the way people usually speak in Pithoragarh. A second was to analyse informal written web data from an online corpus. This will be discussed in the following section.

6.3   Corroborating analysis results using a web-based corpus

Because of the small number of *type* tokens and their subdivision into so many varied uses, the decision was taken to corroborate the results of the spoken data with a readily-available

---

[32] In the personal experience of the researcher

corpus of informal web data: Hindi Web 2012 (HindiWaC v.4).[33] Consisting of 30.6 million words, the corpus is a collection of web pages from 13,321 unique URLs which had been crawled from the internet using search engines between 2012 and 2017. The corpus contains web pages which employ more formal Hindi (such as Wikipedia and news domains like bbc.co.uk, bhaskar.com and jagran.com), as well as interaction-oriented sites such as blogs and health sites where people share their views and post comments in more informal language. A comparative test by the corpus creators found that Hindi Web contains up to five times more personal pronouns than Wikipedia, one of the sub corpora and the site used to generate the initial word frequency list (Kilgarriff et al., 2010). The authors cite this as evidence that the language collected in the corpus as a whole is comparatively more interactional and informal in terms of register. Indeed, a spot-check of the other available Hindi corpus on SketchEngine showed fewer relevant tokens and so the decision was taken to use Hindi Web for its more informal content.

The advent of social media has bridged the gap between spoken and written mediums and created a space where people can write the way they talk; perhaps not exactly, but without the same need for grammatical accuracy and language moderation as in more formal written contexts. Kolyaseva (2018) used posts and comments extracted from an online student forum to research uses of Russian *tip(a)* as a quotative, stating that the posts '[reflect] what can be called spontaneous, everyday speech in writing' (2018:84).

However, whilst it may be interactional, informal and (relatively) unfiltered, written data from the internet is not the same as spoken data. One key area that differentiates the two data sources is that 'interlocutors' are not physically present, nor are responses immediate (as in live chat data like WhatsApp or Facebook Messenger). For these reasons, the Hindi web data will have a limited but important function corroborating the uses of *type* identified in spoken communication.

Contributors to Hindi Web may be from anywhere in North India (where Hindi is most widely spoken), or indeed anywhere else in the world. There is no (easy and legal) way of discovering the identity and location of the writer. This is both a benefit and a drawback: a drawback because the comparison is between a known and an unknown source; a benefit because the majority of contributors are likely to be from outside Kumaon, and because with a corpus of this size there are likely to be several hundreds if not thousands of contributors. There is of

---

[33] The creation of the corpus is described in Kilgarriff et al (2010). Corpus access through the Sketch Engine programme was granted through the ELEXIS programme funded by the EU.

course the possibility that one of the 47 research participants from Pithoragarh may be among the contributors to the online web pages, but this is considered exceedingly improbable for the abovementioned reasons and in light of the population of Hindi speakers globally who could have contributed. The Hindi Web data therefore provides a means of evaluating whether the findings from the spoken data are in evidence more broadly among Hindi speakers/authors.

6.3.1   Extraction of *type* tokens

Since Hindi speakers use both Devanagari and Roman scripts to type informal Hindi on social media, searches were conducted in both scripts. Searching on the string 'type' rendered 946 tokens and all were retained in the sample since an initial perusal of the data showed that many of them would originate from monolingual English sentences and not prove relevant.

Next, a search on the Devanagari string 'टाइप' was conducted, which rendered 3183 tokens.[34]

The picture in Fig 18 illustrates the results of the search. A random sample was generated using a Sketch Engine function. This yielded 320 tokens which were added to the sample.

The 946 tokens of Roman 'type' and 320 tokens of Devanagari टाइप were exported into Excel and formed the triangulation corpus.

Fig 18: Screenshot of Sketch Engine taken on 8th July 2019

---

[34] A search on non-standard spelling 'टाईप' yielded only 478 tokens. A search on both spellings of the noun with plural oblique inflection ('टाइपों' and non-standard 'टाईपों') rendered only 12 and 0 tokens respectively. Plural forms of the noun are also non-standard (see Chapter 8.2.1) and are therefore also not included in the sample.

6.3.2 Numbers of tokens accepted in the sample

All tokens in the sample were analysed and coded according to the language(s) that occurred specifically within the *type* phrase. These were categorised as either English, Hindi, Mixed or Technical, the latter referring to *type* as a key word in a computer programming language, a short form of the compound 'data type'. *Type* phrases were coded as 'Mixed' if they contained both Hindi words and any other English-origin words <u>in addition to *type*</u>, regardless of whether they would be considered loanwords or CS. Phrases consisting entirely of Hindi words alongside *type* were coded as 'Hindi', since *type* has already been defined as an established loanword in the Hindi lexicon. One exception is the example in Fig 19 which contained the string *type of* and was deemed to be an English matrix clause with Hindi loanwords (marked in red), since Hindi appropriation of the loanword *type* in the spoken data always involved use of the Hindi genitive, never the English 'of'. The language mixing in this sentence reveals a lot about the use of English in India and the important role of Hindi loanwords in North Indian English, but does not contribute to this thesis, which investigates uses of *type* in colloquial Hindi.

> <s> Only **such type of** *Sanyasi* & *Netas* can make this country super power. </s>
> 'Only such kinds of holy men and politicians can make this country a superpower.'
> **Fig 19: From bharatswabhimansamachar.in**

The total number of tokens according to the language of the *type* phrase is given in Fig 20. Tokens coded as English or Technical were removed from the sample.

Language of *type* phrase	# Tokens
English	786
Hindi	**450**
Mixed	**16**
Technical	15
Total	1267

Fig 20

The tokens analysed as Hindi and Mixed (N=466) were then analysed for their content and items relating to the verbal use of *type* were removed from the sample; these numbered 214 in total.

Also removed were common uses of *type* in compound nouns such as 'Type A' (diabetes) and 'Roman type' (the font). The use of 'type' purely for translations of Hindi text were also avoided. These were easily identified by brackets (). The breakdown of tokens from each

category is given in the table in Fig 21; note that these numbers include all tokens regardless of the script used.

Use of *type*	# Tokens
Verbal	214
Compound nouns	84
Other non-taxonomic	4
Taxonomic and innovative uses	**150**
Unclear	14
Total	466

Fig 21

The remaining 150 tokens that were identified as taxonomic uses of *type* or innovative extended uses of the taxonomic noun formed the sample that was used for corroboration of the spoken data. These 150 tokens were analysed based on their pragmatic context and syntactic constructions in a similar way to the spoken data (see Section 6.4).

6.3.3   Analysis of script

One way in which the Hindi Web data differs from the spoken data is that the text can be written in either Devanagari or Roman script. A breakdown of the scripts used in the *type* phrase is given in the table in Fig 22.

Script	# Tokens
Devanagari	88
Roman	55
Mixed	7
Total	150

Fig 22

Note that the designation 'Mixed' in Fig 22 relates to the script chosen, not the linguistic origin of the words used. All seven mixed script *type* phrases were written predominantly in Devanagari with only *type* written in Roman script. A more detailed study on the Hindi Web corpus data incorporating an analysis of the ways in which scripts were used and combined is out of scope for this thesis but would be a worthy endeavour.

6.3.4 Tokens of *kind* and *sort*

Since the purpose of the Hindi Web data analysis was to confirm hypotheses developed on the basis of the Pithoragarh spoken data, it was important to also determine whether the unequivocal preference for English lexeme *type* over alternatives *kind* and *sort* would be reflected in the relative frequency of all three lexemes in the Hindi Web data. A similar search was therefore conducted for lexemes *kind* and *sort* with the following results which confirmed that *type* is an established and well-used part of the Hindi lexicon, but *kind* and *sort* are not.

There were a total of 148 tokens of string 'sort' in the Roman alphabet. Of these, 137 tokens were from monolingual English sentences and only 11 tokens were from mixed Hindi-English sentences. There were only 19 tokens of Devanagari transliterated सॉर्ट 'sort' of which 10 tokens were from light verb constructions using the verb *sort karna* 'to sort'. There were 913 tokens of string 'kind', and a total of 12 tokens of Devanagari transliteration in two alternative spellings, काइंड and काइन्ड. The vast majority of *kind* tokens were in monolingual English sentences or clauses, including a few tokens where *kind of* was used as a hedge.

These numbers show that there was an overwhelming prevalence of *type* in comparison to *sort* and *kind* in both the spoken corpus from Pithoragarh and the Hindi Web corpus. Addressing the question as to why *type* would have been borrowed as a categorizer noun in preference to *sort* and *kind* is outside the scope of this thesis, since a satisfactory answer would rely heavily on diachronic data. No further investigation was therefore taken into the uses of *kind* and *sort*.

6.4 Analysis of *type* lexemes

This section outlines the steps and the approach taken to analyse uses of *type* in the primary data set. The approach taken: is based on grammaticalization theory; is focused on context, giving primacy to semantic-pragmatic interpretation; includes the broader context of lexical synonyms; and uses cross-linguistically attested patterns to generate hypotheses and (to a certain extent) also to test them. For each token, the following questions were addressed in order to arrive at a complete understanding of how *type* is being used:

1. What does *type* (or the *type* phrase) mean (or communicate) in this context?
2. How is *type* (or the *type* phrase) structured syntactically?

3. How else could this have been expressed in Hindi and what are the differences in meaning, if any, between the possible options?

The first question relates to the semantic-pragmatic context of use and the second to the syntactic context, the construction in which *type* was utilized. The third question focuses on the identification and analysis of semantically and pragmatically equivalent lexemes in Hindi. This section will briefly outline the approach taken in each of these areas.

Before covering the specific analytical steps taken, mention must be made of the decision to foreground pragmatic context. Because of the (highly syntactic) nature of grammatical change, there could be a tendency to focus predominantly on syntactic changes, for example, the progression of a lexeme from noun to preposition to adverb, and to treat the pragmatic context of the utterance as secondary. This thesis maintains that all communication is contextual, and that understanding context is imperative to understanding the message. In line with Rosenkvist and Skärlund (2013) and others, the approach taken here will blend semantic-pragmatic and syntactic analyses, viewing both as important and prioritising interpretation of discourse context.

This thesis uses the framework of grammaticalization theory to analyse innovative uses of *type* in Hindi. All four parameters of grammaticalization described in the previous chapter have been considered, starting with the use of a linguistic form in a new (semantic-pragmatic) context, the first step in the change process. The analysis therefore proceeded according to the following steps:

1. *Identification of different use cases – firstly semantic-pragmatic and then syntactic*
   There is an expectation that layering will occur and that older and newer uses will exist concurrently
2. *Look for contexts of use in which a new syntactic analysis is the only option*
   A **methodological issue** - how to identify when a change has taken place: 'We will say a rule change has occurred if (a) it has evidently **spread** from the individual and been accepted by the group, and (b) the **constraints** of the former linguistic environment **are no longer obligatory**' (Hopper and Traugott, 2003:48).
3. *Identify possible bridging contexts*
4. *Identify possible paths of grammaticalization*
   Being guided by cross-linguistic pathways. e.g. *Noun > P >* Clause-internal Adverbial > Sentence Adverbial > Discourse Particle (Hopper and Traugott, 2003:37; Traugott, 1995)

The uncovering of specific grammaticalization pathways relies heavily on diachronic data and is (even then) notoriously complex to determine. Perhaps for this reason scholars have

tended not to diagrammatize pathways but rather allude to developments of and within specific constructions (but see Brems and Davidse, 2010, as an exception). Indeed, cognisant of the high degree of interconnection between different constructions and uses, Voghera claims it is not possible to linearize the process of *tipo* grammaticalization in Italian and instead refers to a 'complex net [web] of meanings and functions' (2013:304). After several months of effort it became clear that it would not be feasible to identify clear grammaticalization pathways for Hindi *type*, since there were multiple possible routes that could be hypothesized through the tokens in the data. The attempt was therefore abandoned.

6.4.1 Analysis of semantic-pragmatic meaning

Analysis of functional (pragmatic) context

In order to fully understand the semantic-pragmatic meaning of the lexeme, the context of the utterance was analysed. For the recordings, this entailed listening to segments before and after the utterance in question. From the context, including the speech of all interlocutors, it was possible to identify the speaker's intended meaning. Trusted native speaker contacts played an essential role in the project, both for confirming the content of less audible speech, for discussing alternative lexemes, for clarifying the nuances of meaning and the contexts of use of certain phrases and lexemes, and for testing hypotheses regarding the same. In order to facilitate these discussions, snippets of audio files were on occasion sent to trusted contacts for their input. Some of these trusted contacts were themselves research participants, and wherever possible, the audio file was sent to one of the participants from the recording in question. Where this was not possible, the identity of the speakers was kept strictly anonymous.

The Hindi Web data extract from Sketch Engine provided some text before the segment of text containing *type*, and in many cases the segment of text was long enough to provide an understanding of the context. Where the text provided was too short, and the meaning of the *type* phrase unclear, a portion of text containing *type* (usually the sentence) was entered as a search string on www.google.co.in and the original web page identified. This permitted the reading of a large chunk of text which elucidated the use of *type*. The researcher first analysed the tokens alone and then clarified key findings and hypotheses with a small number of trusted native speakers.

As far as incipient quotative usage is concerned, a broad definition of 'quotative' was followed (see the following chapter), such that any cooccurrence of *type (*or equivalent

lexeme) with a lexical reporting verb (say, speak, etc) and reproduction of thought or speech, either hypothetical or otherwise, was deemed quotative; the aim was to capture an emerging quotative form with as much data as possible to analyse its usage in relation to new quotatives in other languages.

Comparison with alternative lexemes

Cruse (1986:270) asserts that absolute synonyms, i.e. that can be used interchangeably in all contexts, are 'extremely uncommon'. Russian prepositions *tipa* and *po tipu* both developed from *tip* 'type' and are considered 'synonymous' by Kolyaseva (2018), who nonetheless demonstrates that *tipa* is used extensively as a quotative while *po tipu* is less so. The understanding that there will be differences in usage between apparently synonymous lexemes is the justification for the analysis of alternative lexemes. Defining how a lexeme is used vis-à-vis its synonyms contributes to its more complete definition. If centuries old synonyms are seen to display differences in usage, however subtle, then the emergence of a new synonym (a loanword such as *type*) would surely influence the usage patterns of all lexemes. Stanlaw (2004:9) asks pertinent questions relating to this issue in the opening to his volume on the use of English in modern Japanese:

> 'Why is it, for example, that all Japanese people have available to them not only their native Japanese number or colour terms, but also complete sets of basic English terms? Do Japanese and English terms label different categories? Are the English terms used merely as synonyms? Do they carry different symbolic or connotative messages? Are they used in different communicative strategies?... What processes or semantic negotiation, then, are taking place when Japanese people use these forms?'

As discussed in Chapter 4.6, De Smet et al.'s (2018) revised theory of competition predicts that synonymous lexemes will be subject to substitution, differentiation or attraction. Peterson (2017) suggests that it would be 'more fruitful and revealing' (2017:123) to compare the functional use of a borrowing in its source language with that in the recipient language is less enlightening than comparing it with the functions of alternative or (partially-)synonymous forms in the recipient language. Given that *type* seems to have developed extensively in Hindi and comparatively little in English, this thesis follows Peterson's suggestion. How speakers use each of the alternative lexemes will be key to understanding the meanings and usage assigned to *type.* Although a complete study of this issue would require a much larger corpus and preferably diachronic data, the aim in this project was to begin to answer the question posed in Question #3 of the previous section: "How else could this have been expressed in Hindi and what are the differences in meaning, if any, between the possible options?" The investigation took two routes. The first was intuitive and based

on knowledge of the language as well as a brief comparison of all tokens of one lexeme and all tokens of another lexeme. The second route involved data discovery using the software programme Tableau,[35] which was also used to create the charts and graphs used in this thesis.

The process followed for pragmatic analysis is recursive and involves multiple iterations of both allowing the data to 'speak for itself' and revising the analysis in the light of cross-linguistic findings. Categories have been labelled and defined using this combination of techniques. Effort has been made where possible to choose a category label that focuses exclusively on the pragmatic usage and is not tied to any particular syntactic realization. However, in certain cases, a category label containing syntactic details found in prior research was appropriated to more easily facilitate cross-linguistic comparison; one such example is 'Attributive modifier', a term used by De Smedt, Brems and Davidse (2007) and also adopted in Mihatsch (2016) and Kolyaseva and Davidse (2018). The term 'Attributive modifier' in fact encompasses the syntactic status of the type noun as a modifier of adjectives which attribute qualities not to the type noun, but rather to the head noun of the noun phrase. An example from De Smedt et al. (2007) is given in Fig 23.

> Being **an accommodating sort of bloke**, he let me take the car around the paddock at Silverstone (CB – UK magazines)
> **Fig 23: Copied from De Smedt et al. (2007:234, Figure 26)**

The label conveys the fact that the kind of adjectives that appear in these constructions relate to characteristics or attributes, predominantly of people, and subjective evaluation is the defining feature of this function of type nouns. A purely pragmatically oriented label could be 'Evaluative descriptor', although this label would apply to the [adjective + type noun] noun phrase and not solely the type noun. This label is also somewhat opaque and highlights another difficulty in categorizing uses of *type* which are perhaps not as prototypically "pragmatic" in their function. And yet, to persist with the definition of pragmatic maintained throughout this thesis, which relates to <u>meaning in communicative context</u>, the attributive modifier usage is functionally (and pragmatically) distinct from a categorizer function. It is also syntactically distinct from the original categorizer construction.

Another more mundane, practical consideration also influenced the choice of categories for analysing different uses of *type* and relates to the issue of <u>granularity (of detail)</u>. The small

---

[35] Tableau Desktop version 2019.1, a product of Tableau Software, LLC, a Salesforce company.

token size (N=67) meant that creating a proliferation of fine-grained categories would not afford a meaningful analysis as there would be so few tokens in each category. When selecting categories, priority was therefore given to subdividing the more innovative uses of lexemes. The result is a slight loss of detail, particularly in the use of adjectives with *type*: tokens with a special set of adjectives termed by Breban (2010) as 'adjectives of comparison' such as *other*, *different* and *same* could have been separated out into a category Mihatsch (2016) defines as 'Identifying'. However, this did not seem pertinent to the examples in the data set and has not been included.

As is common in academic research, the challenge in choosing category labels and particularly in comparing results from different studies is to ensure equivalence of meaning. In the field of type noun grammaticalization, which overlaps with a great many other fields of research, there are cases where scholars use the same term to refer to (slightly) different phenomena, and other cases where different terms are used to refer to the same phenomenon. An example of the latter is the choice of term employed for the taxonomic usage of type nouns, which is variously labelled 'categorization' and 'classification'; in this thesis, 'categorization' is preferred, as it suggests a broader implementation than the taxonomic classification of life sciences. This issue of the challenges with cross-linguistic comparison will be resumed and discussed more fully in Chapter 10.1.

6.4.2   Analysis of syntactic constructions

The approach followed for syntactic analysis aligns with that taken for pragmatic analysis. Since the assignment of syntactic category is based on the relationship of a word to the other words around it, context proves to be the most significant factor here too. A review of literature was less helpful in analysing the syntax of *type* constructions than in analysing pragmatic meaning, since there was much greater variation in the constructions different languages used. However, at a high level, observations from other languages allowed some predictions to be made and interpreted within the context of Hindi syntax. As with pragmatic usage, knowledge and analysis of equivalent lexemes in Hindi proved instructive, in this case in understanding how *type* was being incorporated into Hindi clauses.

Loanwords are partially defined or identified as such by the extent to which they are morphophonetically integrated into the recipient language. Notwithstanding this generally accepted rule, lexical borrowings have also been assigned a role in transporting structural changes from one language into another. This leaves us with a thorny problem when we

want to assess its syntactic adherence to or departure from the norm: which diagnostics should be used, those which relate to binominal constructions in English, the assumed source of the borrowing, or binominal constructions in Hindi, perhaps using a synonym as a model? One of the commonly used diagnostics for determining the status of the type noun as head of the construction is the relative freedom or restriction with regards to the plurality of the type noun and the category noun. This diagnostic criterion applies to English and also to Russian, where one of the defining characteristics of *type* as a modifier is that the genitive *tipa* 'type' does not usually appear in its plural form *tipov* 'of the types'. Additionally, the combination of plural adjective and singular *tipa* is permitted, as in Fig 24, a phenomenon which Kolyaseva and Davidse consider evidence for grammaticalization (2018:201).

> *Delaetsja popytka ocenit' perspektivy razvitija **sistem razlicnogo tipa***
> 'An attempt is being made to assess the prospects of development of **systems of various sorts**'
> **Fig 24: Copied from Kolyaseva and Davidse (2018:201, Figure 26)**

Since morphological integration is an expected result of lexical borrowing, the decision was taken to analyse use of loanword *type* in Hindi from the standpoint of Hindi syntax, based on the semantically equivalent type nouns *tarah* and *prakaar*. As far as the issue in question is concerned, since neither of the equivalent type nouns typically appear in plural form, the lack of plural *type* in Hindi is to be expected and will be considered evidence of morphological integration of the loanword, not evidence of grammaticalization.[36]

6.5 Analysis of Hindi equivalent lexemes

6.5.1 Identification of Hindi equivalent lexemes

Two main methods were employed for the purpose of identifying alternate words or phrases in Hindi that would be equivalent to *type* in each of the contexts of use: dictionary searches and knowledge of the language.

In their study of anglicisms in the German news magazine *Der Spiegel* researchers Onysko and Winter-Froemel (2011) consulted dictionaries and other reference materials to find the semantic (near-)equivalents of the English lemmas they had identified. They then used a Google search to verify whether the so-called semantic equivalents were used by speakers or not. However, research into grammaticalized use of lexemes in other languages confirms that dictionaries are invariably several years behind spoken language due to an unavoidable

---

[36] It may be argued that the Hindi lexemes themselves have undergone grammaticalization, but the point here is that *type* is following the norm.

time lag in updating the dictionary and do not always capture colloquial usage, most likely a reflection of prescriptivist ideologies. This was the case for Russian *tipo* (Kolyaseva, 2018) and Brazilian Portuguese *tipo assim* (Bittencourt, 1999) whose extended uses are not dictionary attested.[37] Additionally, Salkie (2002) shows that most dictionaries are limited in their ability to provide an accurate picture of how words are used in colloquial speech due to the lack of focus on the (pragmatic) context of use. For this reason, dictionaries cannot be relied upon as a reference for colloquial language and "inside" knowledge of the spoken language and how people use it is essential. The essence of this method of investigation is to ask the question, 'If you didn't say *type* in this utterance, what could you say?' or 'What else could you say in place of *type* that would give the same meaning and the same 'feeling'?' The researcher used both her own knowledge of the local variety of colloquial Hindi to answer these questions and also relied heavily on members of the community, in particular a few individuals with astute linguistic awareness who functioned as trusted advisors.

It should be noted that the process of identifying *type*-equivalents was also an iterative process as analysis and reflection on the uses of alternative Hindi lexemes was instructive in further defining the meanings and uses of *type* and as the meanings of *type* were refined by considering alternative lexemes, new data sets were added to the study.

Dictionaries published by Oxford University Press (OUP) were chosen as the main reference point for dictionary definitions, although others have been consulted (see References). There are two main bilingual dictionaries:

1. Oxford Hindi-English Dictionary (OHED): A comprehensive 1083-page reference dictionary
2. Oxford English-English-Hindi Dictionary (OEEHD): Published with students of English, teachers and translators in mind, the dictionary provides the lexical meanings and Hindi translations of (global) English lexemes

Lexemes which emerged from the process of identifying equivalents of *type* were as follows: *aisa, vaisa, jaisa, sa, tarah, prakaar*. Lexemes which were not chosen and are therefore out of scope of the comparative analysis are other hedges such as *kuch*, and other discourse markers such as *matlab*. Although a full study of these lexemes was not completed, we will

---

[37] Mihatsch also reports that prepositional uses of *tipo* in Argentinian Spanish are 'hardly ever mentioned in dictionaries', although they have been attested since the nineteenth century (2018:164).

see that they (and others) appear alongside *type* and are used to form clusters of vague expressions.

6.5.2   *Type* equivalents data set

All tokens of each of these lexemes were extracted from the corpus via ELAN searches which included retrieval of plural and inflected forms in addition to variant spellings which were cleaned up and standardized. The numbers of tokens retained in the sample for each lexeme are listed in the table in Fig 25. Four tokens were excluded because (part of) the utterance was inaudible which meant that the intended sense of the word could be not accurately determined.

Linguistic form	# Tokens accepted	# Exclusions
aisa[38]	787	1
aise	346	
vaisa	77	
vaise	231	1
jaisa	126	1
jaise	364	
sa[39]	207	1
tarah	65	
prakaar	4	
TOTAL	2207	4

Fig 25

6.5.3   Analysis of *type* equivalents

The process for analysing the pragmatic contexts and syntactic constructions in which each lexeme appears was similar to that followed when analysing *type* and the findings of each instructed hypotheses developed for the other. There is almost no published research on Hindi type nouns or on similatives and similative demonstratives, either regarding the grammaticalization of these lexemes or their pragmatic functions.[40] One exception is

---

[38] This includes the feminine singular form *aisi* and the oblique form *aison*. The same applies to lexemes *vaisa* and *jaisa.*

[39] This includes the feminine singular form *si* but not the plural form *se* which is homophonic with the highly polysemous case marker *se* and could not be easily disambiguated*.*

[40] There is a scarcity of research on pragmatic issues in Hindi in general (and indeed in Urdu, which shares a grammar and a large proportion of the lexicon with Hindi). Some notable exceptions include Butt and Bögel's (2019) work with Biezma and Jabeen entitled 'Urdu/Hindi Questions at the Syntax-Pragmatics-Prosody Interface' and Montaut's (2015) study of the discourse particle तो *to.*

Montaut (1995) on Hindi *jaisa* and *sa*, although the article focuses on similative and intensifying uses, particularly of *sa*, and therefore does not cover hedges. Sahoo and van der Auwera (2019) provide a helpful introduction to similative demonstratives in South Asian languages, but do not consider their more innovative functions. (These two articles will be referenced in Chapter 8 at the relevant points alongside the analysis of Hindi data.) The lack of literature on the subject has meant that this study required pioneering work, and extensive use was therefore made of native speaker intuition and two other methods: comparison of equivalent lexemes and dictionary definitions. Additionally, and similar to the approach taken for loanword *type*, cross-linguistic research on the grammaticalization of type nouns and similative demonstratives was used as a strong reference point for the investigation.

All of the Hindi lexemes appear in the Hindi-English bilingual dictionary OHED, and several of the less grammaticalized uses were confirmed by dictionary attestation, for example, use of *jaisa* as a similative in postpositional construction *ke jaisa* is attested in the dictionary, although its elided version without the postposition *ke* is not. Similar to findings for quotative and exemplification uses of Russian *tipa* (Kolyaseva and Davidse, 2018; Kolyaseva, 2018), more highly grammaticalized forms of the Hindi lexemes including as an adverb and discourse particle and their related semantic-pragmatic meanings were not listed in the dictionary. Various reasons for this were proposed in the previous section. As with *type*, cross-linguistic patterns were used to test hypotheses about the development of the lexemes. Additionally, comparing the lexemes with each other using native speaker intuition (both own and that of advisors) proved instructive in illuminating nuances of meaning and differentiating between the lexemes.

Since the plural form of the similative demonstrative of quality *aise* is homophonous with the similative demonstrative of manner *aise*, the token numbers given in Fig 25 reflect the extracted data not the lexeme, as defined by a combination of form and meaning. An important aspect of analysis involved the differentiation between different meanings assigned to the same linguistic form, as well as the identification of different words fulfilling the same meaning.

6.6 Comparison between *type* and other lexemes

Having identified *type* equivalent lexemes and analysed the semantic-pragmatic contexts and syntactic constructions in which they were attested in the Pithoragarh data, the next

step was to compare the uses of these partially synonymous lexemes one with another. The purpose of the comparison was to differentiate the meanings and uses of each and particularly to aid in the understanding of how *type* has come to fit into the lexicon vis-à-vis its alternatives. This analysis was performed both in Excel and using the data visualisation software programme Tableau.[41] The data set consisting of all *type* tokens and all *type* equivalent tokens was loaded into Tableau from Excel. The programme is flexible and easy to use and facilitated the display of data in charts and graphs in a way that allowed patterns to become visible. This flexible process of "data discovery" allowed the researcher to ask questions of the data and select which variables to analyse and in which combinations. Variables explored included lexical collocation, the part of speech of adjacent words, syntactic constructions and pragmatic functions and more typical sociolinguistic variables such as gender of speaker and topic of discourse. The variables were combined in different ways with the aim of explaining differences or specializations of usage between *type* and its equivalents. In the absence of a principled corpus and an equal number of participants from each sociolinguistic quadrant, it was not possible to conduct statistical tests. Rather descriptive statistics were used to highlight general trends in the data. Chapters 8 and 9 contain the results of this analysis.

6.7   Conclusion

In this chapter the approach to analysing the spoken data and the informal written web data have been described. Spoken conversations between groups of friends, neighbours and relatives yielded 16+ hours of naturalistic data from which all tokens of English-origin lexeme *type* were extracted, along with all tokens of semantically (and pragmatically) equivalent lexemes. The data was analysed using native speaker intuition, dictionary searches and by drawing heavily on cross-linguistic research. A detailed comparison between the lexemes was also key to delineating the uses of *type* in relation to alternative options in the language. The following chapter analyses studies of *type* cognates cross-linguistically and highlights salient points relevant for this thesis. Results from the analysis of *type* and its equivalents follow in the proceeding chapters.

---

[41] Tableau Desktop version 2019.1, a product of Tableau Software, LLC, a Salesforce company.

7 Grammaticalization of *type* cognates and similative demonstratives cross-linguistically

7.1 Introduction to cross-linguistic analysis

The aim of the current chapter is to provide a comprehensive overview of the grammaticalized functions of type nouns and similative demonstratives in the world's languages, with a particular focus on cognates of *type*, and also to highlight gaps in our knowledge, including several areas that will be enriched by this study.

Type nouns, also known as taxonomic nouns or categorizer nouns, are lexemes used for categorizing items within a (pseudo-)taxonomy or hierarchy, e.g. 'A recliner is a *type* of chair'. The lexeme *type* designates a subcategory within a superordinate category, so in the example sentence given, 'recliner' is classified as a subcategory of the category 'chair'. Categorizer nouns cross-linguistically include cognates and synonyms of lexeme *type*, from which the group of nouns earns its name, but also cognates and synonyms of *kind, sort, genus, class, form, (sub)species,* and *style*. One grammaticalization process that has sparked recent interest is the development of type nouns from their use as categorizers to fulfilling such diverse functions as quotatives, hedges and focus markers. Fig 26 displays an example of Russian *tip* functioning as a quotative marker.

> *Snachala nas sprashivali, **tipa**, chto vam nuzhno dlya ucheby...*
> 'At first, we were asked, **sorta**, what do you need for studying...'
> **Fig 26: Reproduced from Kolyaseva (2018:87, example 8, bold mine)**

In recent years, research on type nouns has expanded to include quasi-taxonomic nouns whose 'original' meaning is non-taxonomic, but which seem to grammaticalize in similar ways (i.e. develop the same or similar extended meanings) to taxonomic nouns. These include cognates of *style* (see Kisiel, forthcoming, for Polish *w stylu*) and even Spanish *onda* 'wave' (see Kornfeld, forthcoming).

Each cognate and synonym of *type* has its own grammaticalization history, although a cross-linguistic comparison reveals several similarities between them, both in terms of the functions performed and the process of semantic development towards greater subjective expression (subjectification). The grammaticalization of type nouns cross-linguistically is so extensive that research into the process intersects with research in other more specific areas. One of the most relevant for this thesis is research into a class of lexemes known as similative demonstratives or similarity demonstratives. Intrinsically combining similative comparison with deictic reference, the similative demonstrative is prolific in South Asian languages,

including Hindi, and coexists with loanword *type* in the Hindi lexicon. However, similative demonstratives exist in many language families and are as yet under researched, even more so than type nouns whose grammaticalization has been the subject of recent focus. Many of the functions performed by grammaticalized type nouns are also extended functions of similative demonstratives, as shown in Fig 27, which illustrates the use of German *so* 'like this/that' as a quotative.

> *Und ich* **so** *"…". Und er dann* **so**: *"…"*
> 'And I'm **like** … And he's **like** …'
> **Fig 27: Copied from König (2017:160)**

In section 7.3 of this chapter, the category of similative demonstrative is defined and distinguished from 'ordinary' demonstratives, and research into the grammaticalization of similative demonstratives cross-linguistically is presented.

Finally, Section 7.4 focuses on interrelationships between type nouns and similative demonstratives that have been observed in several languages. Whereas in some languages the two are seen to frequently combine, such as in Brazilian Portuguese where *tipo* 'type' and *assim* 'like this/that' have ultimately combined to form a compound, in Norwegian we note that both appear in similar semantic-pragmatic contexts fulfilling similar functions and that the development of *type* seems to have been influenced by that of *sånn*. De Smet et al.'s (2018) research on the principle of attraction in language change is explored as an explanation for this latter phenomenon.

7.2 Grammaticalization of *type* cognates cross-linguistically

A cross-linguistic comparison of the grammaticalization of *type* reveals a great degree of uniformity at a high level and an equal amount of variation at the detailed level, not only between languages, but also between specific constructions within a language. Both Italian (Voghera, 2013) and Argentinian Spanish (Mihatsch, 2018) are examples of this, to name but two languages. All of the languages surveyed have *type* as a loanword either directly from (Ancient Greek via) Latin or indirectly via another language. The cross-linguistic comparison presented here, in conjunction with that of similative demonstratives, will form the basis for understanding the development of *type* in Hindi.

7.2.1 Overview of the semantic-pragmatic functions of *type* cross-linguistically

Table 1 provides an overview of how cognates of *type* have grammaticalized cross-linguistically by focusing on their various semantic-pragmatic functions. The table has been completed with reference to major works on the grammaticalization of *type* cognates.

Ticks designate attestation in the source literature which has been given below. A cross designates that the type noun is not used to fulfil that particular semantic-pragmatic function. A blank cell indicates that the usage is not mentioned in the source document; this may mean either that the usage occurs but has not been specified by the author(s), or that the usage is not attested. A tick surrounded by brackets (✓) indicates that the type noun has been attested in the relevant context but is not commonly used as such. The functions are arranged from left to right following an approximate order from least grammaticalized to most grammaticalized, but not denoting a linear pathway, as discussed in Chapter 6.4. In certain cases, this can be understood as an implicational hierarchy; for example, a *type* cognate will never grammaticalize into a focus marker or filler without first having developed a hedging function.

Since the number of meanings and functions developed by *type* cognates and synonyms is extensive, I have restricted the columns in the chart to those semantic-pragmatic functions which are most pertinent to this thesis, namely, those contexts in which *type* has been attested in the Pithoragarh Hindi data.[42] Partly due to space limitation, and partly because 'rounders' are not attested in the Pithoragarh data, the different kinds of hedging have been grouped together in the chart and not disambiguated. A column titled 'Other' is used to acknowledge additional non-nominal functions of *type* which were not identified in the Pithoragarh data but have been attested elsewhere. Similarities and differences between the apparent development of *type* in Hindi vis-à-vis cognates and synonyms of *type* in other languages will be discussed in Chapter 10: Discussion.

Source material for cross-linguistic overview of type functions
Following is a list of the sources whose contents have contributed to the compilation of Table 1. Priority was given to languages for which a full-length study has been completed, which

---

[42] In many languages, including Swedish, Italian and English, type has a nominal sense 'model', which in some cases predated the categorizer sense. Italian additionally contains several nominal uses of *tipo* with different meanings as a result of semantic shift: 'standard', 'exemplar', and 'characters for printing' (also borrowed into English) and a grammaticalized use as an adjective meaning 'ideal' (see Voghera, 2013)

excludes languages that also display highly grammaticalized uses of *type* cognates, such as (Modern) Greek *typos/tipos*, but for which research has not yet been completed.[43]

Germanic languages

English *type, kind* and *sort*: Cheshire (2007); De Smedt et al. (2007); Davidse, Brems and De Smedt (2008)
Swedish *typ*: Odden (2019); Rosenkvist and Skärlund (2013)
Norwegian *type* and *typ*: Odden (2019)
German *Typ*: Umbach (forthcoming)

Romance languages

French *type*: Mihatsch (2016); Vassiliadou et al. (forthcoming)
Italian *tipo*: Voghera (2013); Mihatsch (2016)
European Spanish *tipo*: Mihatsch (2016, 2020)
Argentinian Spanish *tipo*: Mihatsch (2018)
European Portuguese *tipo*: Mihatsch (2016, 2020); Marques (forthcoming)
Brazilian Portuguese *tipo*: Bittencourt (1999)

See also Mihatsch (forthcoming b) for an overview.

Slavic languages

Russian *tip/tipa/po tipu*: Kolyaseva and Davidse (2018); Kolyaseva (2018); Kisiel and Kolyaseva (forthcoming)
Polish *typ/typu/w typie*: Kisiel and Kolyaseva (forthcoming)
Czech *typ/typu*: Janebová, Martinková and Gast (forthcoming)
Latvian *tipa/kipa*: Vassiliadou et al. (forthcoming)

---

[43] Preliminary research by Vassiliadou (p.c.) identifies hedging, general extender and quotative uses of Greek τύπος *typos* (alternate romanization: *tipos*). Below are a couple of examples:
Hedging:- Είδα ένα σκυλί λαγωνικού **τύπου** *Eída éna skylí lagonikoú **týpou*** 'I saw a **kind of** (a) beagle'. Note that this could have a categorizing or an approximative (hedging) meaning, but either way it is subjective, highlighting the speaker's point of view, i.e. the speaker is not sure if the dog was actually a beagle.
Quotative:- Αν συμβεί οτιδήποτε τέτοιο, πες **κάτι τύπου** "πραγματικά εκπλήσσομαι με την αντίδρασή σου."
*An symveí otidípote tétoio, pes **káti týpou** "pragmatiká ekplíssomai me tin antídrasí sou.*"
'If anything like that happens, say **something like**, "I'm really surprised at your reaction."'

Language	Word	Categorizer	Phoric modifier	Attributive modifier	Ad hoc categorizer	General extender	Similative	Exemplifier	Hedge	Quotative	Focus marker	Filler	Other discourse pragmatic
English	*type*	✓	✓	✓	✓	(✓)			✗	✗	✗		
German	*Typ*	✓							✗				
Norwegian	*type*	✓	✓	✗	✓	✓		(✓)	(✓)	(✓)	(✓)		
Norwegian	*typ*	✗	✗	✗	✗	✗		✓	✓	✓	(✓)		✓
Swedish	*typ*	✓			(✓)		[44]	✓	✓	✓	(✓)		✓
French	*type*	✓	✓	✓		✗		✓	(✓)				
Italian	*tipo*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
European Portuguese	*tipo*	✓	✓	✓		✗	✓	✓	✓	✓	✓	✓	✓
Brazilian Portuguese	*tipo*	✓			✓		✓	✓	✓	✓	✓		✓
Peninsular Spanish	*tipo*	✓	✓	✓	✓	✓	✓	[45]	✓[46]	✓	(✓)		✓
Argentinian Spanish	*tipo*	✓	✓[47]		✓	✓[48]	✓	✓	✓	✓	✓	✓	✓
Russian	*tip*[49]	✓			✓	✓[50]	✓	✓	✓	✓		✓	✓
Polish	*typ*	✓			✓	[51]	✓	✓	✓	✓	✗		

---

[44] Although many of the examples in Rosenkvist and Skärlund (2013) can be interpreted as similative uses, they can also all be interpreted as exemplifying uses, and the authors seem to prefer this approach.

[45] Mihatsch (2020) notes a single token in C-ORAL-ROM which can be analysed as a bridging context between a similative and an exemplifier.

[46] According to Mihatsch (2020:155), Spanish young people accepted rounder use but not mitigator use of *tipo*.

[47] Mihatsch, p.c.

[48] Mihatsch, p.c.

[49] A smaller subset of the functions listed here are also fulfilled by *po tipu*, whose functions are more restricted than those of *tip(a).*

[50] Alena Kolyaseva (p.c.) has provided an example of general extender usage: *(ili) tipa togo* '(or) *tipa* that-GEN.n.sg' 'or something like that'.

[51] This function is mentioned only in relation to Polish *rodzaj* 'kind', not *typ* 'type'.

Language	Word	Categorizer	Phoric modifier	Attributive modifier	Ad hoc categorizer	General extender	Similative	Exemplifier	Hedge	Quotative	Focus marker	Filler	Other discourse pragmatic
Latvian	*tipa/ kipa*	✓			✓		✓		✓				✓
Czech	*typ*	✓			✓		✓	✓	✗	✓			

Table 1: Cross-linguistic comparison of semantic-pragmatic functions fulfilled by *type* cognates

7.2.2 Functions of *type* cognates: categorizer and manner descriptor

Categorizer

Although type nouns have an etymology that precedes their use in categorization, the function is considered the foundation for the process of grammaticalization under study in this thesis, and common to all type nouns, including all *type* cognates. As explained in the Introduction, categorization involves reference to sub-categories of a named category, such as the example in Fig 28 where *Pneumocystis carinii pneumonia* is described as a (rare) subcategory of category *infection*. The lexeme *type* identifies the subcategory.

> In 1981 a doctor found 5 previously healthy young men with Pneumocystis carinii pneumonia, **a very rare type of infection**. (CB – UK ephemera)
> **Fig 28: Copied from De Smedt et al. (2007:232)**

Adjectives, such as *rare* in Fig 28, modify the subcategory (*rare type*) and not the category, i.e. *rare infection* (De Smedt et al., 2007).

Manner descriptor

Evidence of nominal polysemy involving a taxonomic sense and a manner 'way of doing' sense is well documented. Indeed, there is evidence from etymological research that type nouns have developed from manner nouns, such as French *manière* and its etymological equivalents whose taxonomic meaning developed and was subsequently lost (Mihatsch, forthcoming a). The earliest recorded uses of *manner* in the English of the middle ages include both tokens with a 'type' meaning and with a 'manner' meaning, both of which were in circulation of the Old French of the period from which *manner* was borrowed (OED; see Fig 29 for a clear use of *manner* as a taxonomic noun from a document dated pre-1225).

> *Crabbe is an manere of fissce in þere sea*
> **Fig 29**

Mihatsch provides a cognitive explanation for the high degree of polysemy in lexemes which mean 'type' and 'manner' in terms of a 'conceptual permeability between… subcategories of nominal concepts and verbal concepts' (forthcoming a). It is therefore entirely plausible that semantic shift could occur in the opposite direction (from type to manner meaning). Despite this, there are few current examples of cognates of *type* fulfilling a manner meaning as well as a categorizer meaning in synchronic data. However, there are several examples of this polysemy from other type nouns, including German *Art* 'kind', 'manner' (Umbach, forthcoming), as in Fig 30.

> ***Die Art**, wie sie sich mit ihm unterhielt, erinnerte mich an frühere Zeiten.*
> **The way** she talked to him reminded me of earlier times.
> **Fig 30: Example from Umbach (forthcoming, example 19a)**

Other examples include Czech *druh* 'manner', 'sort' (Janebová et al., forthcoming) and Portuguese and Spanish *forma* 'kind', 'manner, way' (Mihatsch, 2016). In many cases, the type noun forms an adverbial phrase, such as in Russian where adpositional phrase *v ètom rode* 'in this kind' contrasts with adverbial *v takom rode* 'in such a way' (Kisiel and Kolyaseva, forthcoming).

7.2.3   Functions of *type* cognates: phoric modifier and phoric similative

Phoric modifier

One of the oldest grammaticalizations of taxonomic nouns is their usage as phoric modifiers in (post)determiner phrases meaning 'this/that type', such as the example from Italian in Fig 32.[52] Grammaticalized (post)determiner phrases are attested sometime after the taxonomic meaning of a type noun first appears in a language; in the case of English *type* this occurs almost immediately (Mihatsch, forthcoming a). The demonstrative aspect of the phoric modifier construction, pointing to a specific object or person in the physical context (exophoric) or the discourse context (endophoric), readily produces an 'instance' reading rather than a 'kind' reading. In Fig 31, for example, the speaker clearly wishes to comment on the softness of a particular bag. Using a type noun allows him/her to make a broader statement situating the bag in question within a subcategory of soft bags, but the focus is on the specific bag in view (or under discussion), not the kind of bag it is.

> *allora **quel tipo di borsa** / è necessariamente morbida* [C-ORAL-ROM ifamdl04]
> 'Now **that type of bag** is necessarily soft'
> **Fig 31: Copied from Mihatsch (2016:153, example 70, translation mine)**

Another pragmatic feature of the phoric modifier construction with *type* is that of pseudo-scientific or pseudo-academic connotation, where the speaker can reference a subcategory thus adding weight to their argument in discourse (Mihatsch, 2016). This aspect of *type* usage perhaps arises from its etymological status of <u>not</u> forming part of scientific taxonomies but rather being introduced as the Everyman's categorizer.

Connected to the phoric modifier expressions with *type* are other postdeterminer constructions that Mihatsch (2016, forthcoming a) subdivides into 'identifier' and 'quantifier'

---

[52] Although the only examples in the literature involve postdeterminer phrases, syntactic phrases involving another word order but achieving the same semantic-pragmatic effect in other languages must be included in order to ensure true cross-linguistic comparability.

expressions. Identifier constructions utilize non-demonstrative determiners such as *another* and (adjective) *different* which, due to the element of comparison inherent in their semantics, also entail phoric reference. Quantifier constructions are formed of a quantifier such as *all* or *every* in combination with *type* and produce an expression whose semantic sense highlights the individual subcategories. Both identifier and quantifier expressions thereby promote instance readings rather than kind readings. Since these have not been identified in the Hindi data, they have not been focused on in this thesis.

Davidse et al. (2008:151) describe the 'discourse function' of the phoric modifier as 'anaphoric and analogic to *such*'. Semantically, along with a category reference ('this/that type of N') it also incorporates a similarity element ('an N like this'). This similative connotation is not as pronounced as in the phoric similative construction, where it is unequivocal. However, by focusing on the instance being referred to, either in the physical word or in discourse, a comparison is drawn between the item in question and other members of the named category which are evoked by the construction, i.e. in the example in Fig 31 above, other bags similar to the one in focus.

Phoric similative

Whereas through the phoric modifier construction the speaker intentionally draws attention to the category whilst highlighting the individual object under discussion, the phoric similative is inherently vague, incorporating only the deictic demonstrative and not the category noun, which is neither present in the *type* phrase nor accessible anaphorically. The result is a vague and deliberately under-specified communication relating to *this/that type*, but without specifying of what the object under discussion is a type. The interlocutor is left to infer a relevant category, relying on discourse context and shared knowledge. In Fig 32, the category can be understood as 'men', or, more broadly, 'people'.

> *He's just **that sort**. / He's just **like that**.*
> **Fig 32**

Without a category reference the construction cannot categorize, but instead evokes a similarity relation of *like this/that*. The similative reference is equally vague and the hearer must use discourse context and shared knowledge to extrapolate which elements discussed (if any) are salient items for comparison. In more innovative uses of the phoric similative, similative comparison is impossible and vague connotations are evoked, often negative, as could be construed from Fig 32, although a positive reading is also possible, depending on the discourse context.

Among the full-length studies surveyed, only Norwegian shows a type noun functioning in any way close to the phoric similative, where a fixed 'chunk' comprising the demonstrative and type noun combined can appear without a following category noun as the subject of "presentative" sentences (Odden, 2019:160). The native Norwegian lexeme *slags* (which developed a taxonomic meaning in the sixteenth century) shows evidence of being highly grammaticalized and is commonly found in this context, but there are also examples with *den type(n)* 'that type', such as in Fig 33.

> *Jeg vet det finnes* **den typen**. *De irriterer meg grenseløst.*
> I know **that type** exists. They annoy me endlessly.
> **Fig 33: Copied from Odden (2019:180, example 93, translation mine)**

Due to differences between Norwegian and English, it is not possible to translate *det finnes den typen* with an English similative demonstrative lexeme ('such') or phrase ('like that'); however, in Norwegian, *den typen* is substitutable with (pronominal) similative demonstrative *sånn* 'like this' in this context. For further discussion on this critical issue, see the following sections of the chapter on similative demonstratives and the relationship between type nouns and similative demonstratives.

7.2.4   Functions of *type* cognates: attributive modifier and ad hoc categorizer

Attributive modifier

The attributive modifier is a binominal construction comprising an adjective, the type noun and the category noun, as in *an accommodating sort of bloke* (De Smedt et al., 2007:235; see also Davidse et al., 2008). The adjective is often an 'unusual lexical item' (De Smedt et al., 2007:238) and the pragmatic effect created by the construction is an evaluative one, illustrating increased subjectification, one of the features of grammaticalization outlined in Chapter 5.2. In contrast to the categorizer construction, which appears formally similar, the adjective refers not to the type noun but modifies the category noun itself; so in the example above, the hearer is meant to interpret that the *bloke* is *accommodating*. Semantically, the reference is therefore to an instance and not a type (subcategory). Syntactically, this triggers a 'head reversal', in which the type noun ceases to be the head of the noun phrase, surrendering this role to the category noun. In both of these respects, the attributive modifier is similar to the phoric modifier.

The fact that seminal research on type nouns (in English) included this construction may have given the impression that it is more significant cross-linguistically than it is, as Table 1 illustrates. In her comprehensive overview of type nouns in Germanic, Romance and Balto-Slavic languages, Mihatsch (forthcoming a) claims that 'modifier uses seem less entrenched'

possibly due to their preference for postposed type noun phrases which may have blocked further development. In English itself, it only accounts for 2.5% of type noun uses (Davidse et al., 2008:159). Certainly, in terms of universals, the phoric modifier is much more of a contender for universal status. Romance languages attest infrequent attributive modifier uses, would not usually render an approximative interpretation, although there are examples (such as Fig 34) which show semantic ambiguity and could equally be interpreted as hedges (Mihatsch, 2016):

> *Por lo demás, Quincy Jones ha sido* **un bien curioso tipo de coproductor,** *…*
> For the rest, Quincy Jones has been a very curious kind of co-producer, …
> **Fig 34: Copied from Mihatsch (2016:150, Fig 51, translation mine)**

Uses are also attested with *type* cognates in Romance languages, but seem to be much more frequent with cognates of *form* than with other type nouns (Mihatsch, 2016:156-7). The reasons why Romance languages and Russian do not develop attributive modifier uses are not totally clear, but Mihatsch (forthcoming a) suggests it may be because their preference for postposed adjectives (postpositional variants) has blocked further development.

Ad hoc categorizer

Ad hoc categorization is another common function of type nouns in general and *type* cognates in particular. Ad hoc categories are created on-the-fly by speakers to enable them to describe an unconventional class of objects or people on the basis of a named example or descriptive text or evocative adjective placed 'under a label of a well-established category' (Mihatsch, 2018:153).

> *En Extremadura* **el concejo colectivista del tipo de Sayago** *debe ser muy frecuente.*
> 'In Extremadura **the Sayago-type collectivist council** must be very frequent' (lit. collectivist council of the type of Sayago)
> **Fig 35: Copied from Mihatsch (2018:161 – Peninsular Spanish example from 1898)**

Fig 35 is an example from Peninsular Spanish. In all the languages studied, the typical construction is "X of the type of Y" and is formed of three parts: X, the type noun, and Y, where X and Y are both nouns or noun phrases.
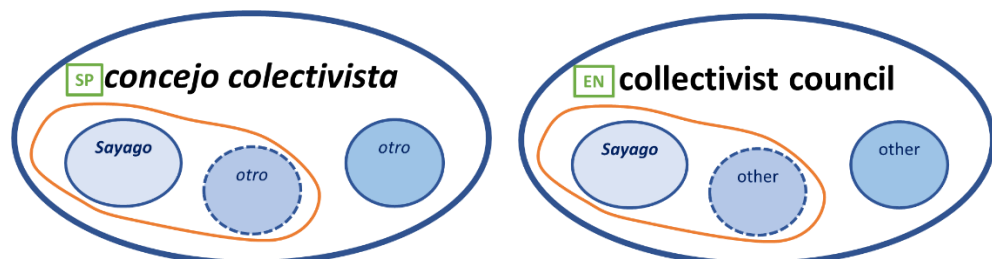
Fig 36: Visual depiction of the ad hoc (sub)category cognitively generated by the phrase *concejo colectivista del tipo de Sayago* **(with translation)**

This kind of classification conjures up a single class of such objects or people and does not refer to a hierarchy (see Fig 36 for a visual depiction). This usage of *type* cognates is extremely prolific in Romance languages and particularly with *type* whose etymology from a prototype, model or exemplar meaning renders it well-suited to connect objects at the same taxonomic level, rather than assign them to a subcategory of a named category (see Mihatsch (2018) for a detailed explanation).

Although English and German do not have a trinominal construction (at least not in the "X + TYPE + Y" format), perhaps one reason why type nouns do not develop a similative function in these languages, English does have a semi-suffix construction which can be considered the translation equivalent of the trinominal construction in Romance languages (Mihatsch, 2018). The English semi-suffix consists of a descriptive element preceding a (cliticized) type noun followed by the category noun, e.g. *a keeping-up-with-the-Joneses-sort of person* (Davidse et al., 2008:161), in which an ad hoc subcategory of people is created. Odden (2019:187) claims the English construction has been the model for a similar use of Norwegian *type*, as illustrated in Fig 37. The semi-suffix construction is infrequent in Norwegian and quotation marks around the LH element are expected (when in written form), suggesting that the construction is more marked in Norwegian than in English.

> Ikke akkurat "første date" type mat.
> 'Not exactly "first date" type food'
> **Fig 37: Copied from Odden (2019, example 97b)**

Regardless of the language or construction, in all cases the ad hoc categorizer usage of *type* creates an instance not a type reading: in Fig 38 the focus is on the food encountered during the first date. It is common for the adpositions (meaning 'of') to be elided and for the resulting construction to resemble "X + TYPE + Y", further developing into the approximative "TYPE + Y".

At a detailed level, there are further specificities in many languages, whether semantic or syntactic: in Czech construction "*X typu Y*", X is always a more general noun whilst Y is a more specific one (Janebová et al., forthcoming), and in Polish, which has a similar construction, the noun in the right-hand position must be plural (Kisiel and Kolyaseva, forthcoming).

7.2.5   Functions of *type* cognates: general extender

A general extender functions as an ending to an utterance which extends the content of the utterance, rendering it broader and more open-ended and according to Aijmer (2002:13) involving 'an element of vagueness in the proposition'. Although the alternative terms 'list completer' and 'set-marking tag' suggest that general extenders such as 'and stuff' and 'or

something' complete and otherwise incomplete list of items in a set, Traugott defines general extenders as '[extending] the prior text rather than the list' and playing a 'metatextual' role in discourse (2016:29). In fact, Pichler and Levey's (2010) research on the development of general extenders suggests they progress from belonging to the end of a set of items to functioning as a topic-closing device and a hedge. General extenders can take the form of a vague noun, as in Norwegian *og ting* 'and things' (Andersen, 2010) and NZ English 'and that kind of thing'; 'and things like that' (Terraschke and Holmes, 2007). General extenders also often incorporate a similative, such as the following example (Fig 38) from Overstreet (2014):

> I thought some poison in his drink **or something like that** a few weeks down the road would kill him
> **Fig 38: Copied from Overstreet (2014:109)**

Uses of determiner complexes with a meaning similar to 'that type of thing' are common in Romance languages (Mihatsch, 2016:153-154) and can have a list completer or an anaphoric deictic reference. One such example is from Spanish (referenced in Mihatsch 2016:154): *este tipo de cosas* 'that kind of thing' (lit. 'this type of things').[53] In English, general extenders with *kind* and *sort* tend to appear in studies, whereas mention of *type* is rare: only a single general extender with *type* ('and all this **type** of stuff') was attested in a corpus of speech of British adolescents described in Cheshire (2007). Overstreet (2014) comments on the general extender 'and that type of thing', referencing Dines (1980), and surmises that '[p]erhaps **due to its infrequency**, this construction has received little attention in studies on general extenders' (Overstreet, 2014:108, **bold** mine). The comparison between general extenders in different varieties of English by Aijmer (2013) likewise details several general extenders involving *kind* and *sort*, but none with *type*. This underuse of *type* compared to *kind* and *sort* in English perhaps reflects the relative level of grammaticalization of the three lexemes. Aijmer (forthcoming) classifies 'sort of thing' as a discourse marker rather than a general extender and describes mitigation, intersubjective and ironic uses. This multiplicity of analyses reveals the multifunctional nature of these phrases and indeed of type nouns in general.

---

[53] General extenders in English use *sort and kind*, which are more highly grammaticalized than *type*, and also tend to employ the distal demonstrative 'that' rather than 'this'; see Terraschke and Holmes (2007) for a comprehensive list of examples from New Zealand English.

7.2.6 Functions of *type* cognates: similative and exemplifier

Similative

Similative comparison describes a similarity between two entities that resemble each other and is expressed through different strategies in the world's languages, from nouns to prepositions to verbs to affixal morphemes (König, 2017; Treis and Vanhove, 2017). In English, similative comparison is expressed by *like*, as in 'Mark is (tall and thin) *like* a giraffe', but also the complex preposition *similar to,* and (less commonly) the verb *resemble*. Similatives can be distinguished from what are usually referred to as comparative constructions, which express similarity or difference between objects by degree, establishing a relationship of inequality (as in 'Dinesh is *older than* his brother'). In many (but not all) languages, they are also distinct from equative expressions such as the English construction *as… as*, which communicate that two objects are alike in equal degree.

Languages differ in how they realize these three kinds of comparison; indeed, not all languages have a separate similative marker and different kinds of comparison may be realized by the same lexeme or morpheme in many languages (Fortescue, 2010).[54]

The seminal cross-linguistic study by Haspelmath and Buchholz on equatives and similatives takes the premise that 'similatives express identity of manner' (1998:313) and thus focuses almost exclusively on manner similatives (e.g. 'She walks like an elephant'), ignoring the similarity of quality ('Your dress is like mine') almost entirely. Treis and Vanhove's (2017) volume on the grammaticalization of similatives and equatives covers an impressive range of language families and details cross-linguistic grammaticalization of similatives as a hedge and approximation marker, quotative and purpose clause marker among other functions. However, it too is also primarily concerned with manner similatives. Motivated by data, this study argues for a broader definition of similatives to include a subdivision between similatives of manner and similatives of quality (like kind).

The development of *type* as a marker of similative comparison is a feature of *type* grammaticalization in all the language families studied with the exception of Swedish, which seems to have developed only the exemplifier usage (see below), although these are closely-related. Voghera (2013) proposes that one reason Italian *tipo* has developed a function as a

---

[54] See Treis and Vanhove's excellent (2017) volume which is focused on similatives and equatives, but contains many chapters also covering comparatives.

similative marker is that Italian *come* 'like, as' is used for both equative and similative comparison. An example of *tipo* functioning as a similative is shown in Fig 39.

> *Non fare **tipo lui***
> 'Don't do **like him**!'
> **Fig 39: Copied from Voghera (2013:284, Figure 3)**

Syntactically, when grammaticalized *type* functions as a similative, it is no longer nominal, but rather adpositional (or forming part of an adpositional phrase). This change of syntactic category (decategorialization) is evidence of grammaticalization.

Exemplifier

Closely related to the similative both semantically and syntactically is the exemplifier function. Often translated as 'for example', 'for instance' or 'such as', but also more colloquially as 'like', the role of the exemplification marker is to introduce an example which illustrates the point the speaker has been making. In Italian and Russian, the exemplifier seems to have developed out of the similative and the strong cognitive connection between the two is apparent; whereas the similative compares one object or concept to another, the exemplifier presents an object or concept as one example (among other possible examples) of something similar to the object or concept just described. In Fig 40 from French, *commode* and *buffet* are presented as two possible examples of *meuble de rangement* 'storage unit'.

> *… on peut envisager une séparation par **un meuble de rangement, <u>type</u> commode ou buffet**.*
> '… one can consider separation by **a storage unit, <u>such as</u> a chest of drawers or a dresser**'
> **Fig 40: Copied from Vassiliadou et al. (forthcoming)**

The Polish case is curious: Polish *typu* and *w typie* have both developed a different syntactic construction from the prepositional similarity marker and yet do not seem to fulfil the requirements of a particle in the same way as other Polish particles. The exact categorial nature of *typu/w typie* in this construction is therefore still under deliberation by scholars (Anna Kisiel, p.c.); it seems to be a construction in its own right. Furthermore, *w typie* can only be used for quotation, whereas *typu* can be used for other forms of exemplification, such as in Fig 41:

> *inne metody **typu** hipnoza, regressing*
> 'other methods **such as** hypnosis, regressing, (…)'
> **Fig 41: Copied from Kisiel and Kolyaseva (forthcoming, example 48)**

Hedge

A linguistic hedge is a word or phrase employed by the speaker to indicate to the listener that what they are communicating is in some way an approximation or to minimise the force of the utterance. Various attempts have been made to subdivide hedges into different kinds. See Kaltenböck, Mihatsch and Schneider (2010) for a helpful and comprehensive summary of the major early works on hedging, which illustrates the various terms employed by scholars to subcategorize hedges according to their different types. One of the earliest and most influential works on hedging was Prince et al. (1982), summarised in Mihatsch (2010). Prince et al. identified three kinds of linguistic hedges: 'adaptors', which operate on the lexical level and communicate that the word chosen is not exactly right; 'rounders' express numerical approximation; and 'shields' attenuate any negative consequences of the utterance by softening the illocutionary force of a speech act, such as 'mitigating a piece of advice' in order to avoid appearing 'patronizing' (see Mihatsch, 2010:111). The seminal work on linguistic politeness by Brown and Levinson (1987) covers this shielding use of hedges for politeness or face-saving reasons and lists numerous hedges in Tamil, English and Tseltal.

An example from Argentinian Spanish can illustrate the function of a hedge. In Fig 42, the use of Spanish *tipo* placed directly before the word *charlas* 'talks' communicates a sense of approximation, that what 'they' wanted to have were not exactly talks, or that 'talks' was not quite the right word for it. *tipo charlas* thus means 'something like talks', indicative of the strong similarity reading which gives rise to the approximative reading.

> *querian hacer **tipo** charlas*
> 'They wanted to have **[something] like** talks'
> **Fig 42: Copied from Mihatsch (2018:166, Example 22)**

Likewise, in Fig 43, Swedish *typ* is a hedge on the phrase *inga pengar* 'no money', which flags to the hearer that the phrase is less than completely accurate, i.e. the speaker has some money and is likely using hyperbole.

> *Just nu har jag **typ** inga pengar*
> 'Right now I **like/kind of** don't have any money'
> **Fig 43: Copied from Rosenkvist and Skärlund (2013:324)**

Mihatsch's (2007) study of hedges in Romance languages reveals an implicational relationship between the use of a lexeme as a hedge for lexical approximation and its subsequent usage as a hedge for quantification ('rounder'). The semantic map in Fig 44 indicates the direction of grammaticalization using arrows.
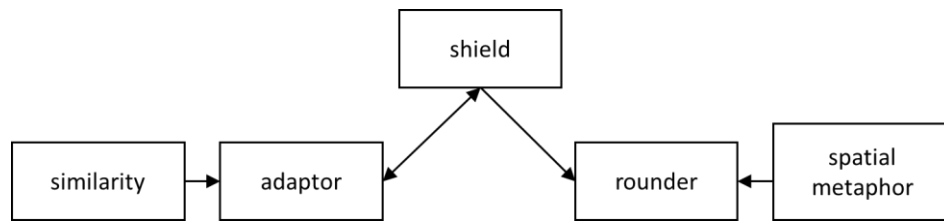
Fig 44: Implicational semantic map copied from Mihatsch (2010:117, Figure 5.1)

Mihatsch found that a lexeme will only be attested as a rounder if it has first been attested as a shield (used for mitigation purposes). The data studied accords with one of the universal principles of grammaticalization, unidirectionality; not every type noun which had developed a lexical approximative function also performed an quantifying approximative function, but the reverse was never true.

In the chart, lexical approximation has not been separated from quantification but both are subsumed under the category "Hedging for approximation", since rounder usage is not (yet) of great import in the Hindi data. Hedging uses of *type* are one of the most common developments and can be seen as non-negotiable within the general type noun pathway; any type noun progressing to discourse marker functions will always have developed hedging functions first. We note from the chart that less grammaticalized type nouns such as Czech *typu* have developed functions as a similative, exemplifier and quotative but not as a hedge.

Regarding the syntactic status of hedges, it must be noted that lexemes from any grammatical category can function as a hedge, as Fraser aptly asserts: 'an expression is usually only recognized as a hedge when it is used in hedging… hedging devices are drawn from every syntactic category.' (Fraser, 2010:23) In the context of type noun grammaticalization, hedges are realized within two primary syntactic contexts: binominal constructions known as 'nominal qualifiers' and grammaticalized adverb or discourse particle uses of the type noun; in Romance languages, these have developed from trinominal type noun constructions (see Mihatsch, forthcoming b).

There is some variation within languages as to which derivatives (or which constructions) of type noun can be used as hedges and which not. In Polish, preposition *w typie* can be used as a hedge (or what Kisiel and Kolyaseva (forthcoming) term a 'vagueness marker'), whereas *typu* does not carry such approximative meaning. This is, in fact, part of a productive development within Polish in which complex prepositional phrases comprised of a noun and a preposition have been developing into hedges in recent years (see the recent chapter by Kisiel, forthcoming).

7.2.8   Functions of *type* cognates: quotative

One common function of grammaticalized type nouns which has become a recent research interest is the marking of quoted speech, both real and hypothetical. Almost all of the *type* cognates surveyed have developed the quotative function (see Table 1). In some languages *type* is used alongside a reporting verb, as in Fig 45, which shows an example from Russian.

> *Snachala nas* **sprashivali, tipa***, chto vam nuzhno dlya ucheby...*
> 'At first, we were **asked, sorta**, what do you need for studying...'
> **Fig 45: Reproduced from Kolyaseva (2018:87, example 8)**

There are two pathways attested in the literature: hedge to quotative and exemplifer to quotative. Both are cognitively motivated. Quotatives develop from hedges on the basis that the quotation is only an approximation of what was actually uttered. Mihatsch has also identified a common pathway from exemplifier to quotative functions, based on the cognitive perception of reported speech as an example of what someone might say, or might have said (see Mihatsch, 2007:239 on Brazilian Portuguese and forthcoming (a) for a more general appraisal).

While investigating the quotative *be like* in English, Buchstaller (2014) discovered that *be like* was first used to mark internal thought before developing as marker for reported speech. It is possible that the same pathway applies to *type* cognates and other type nouns cross-linguistically; more research in this area is required.

As with each of the functions developed by *type* cognates, there are differences in the ways that the quotative is realized in different languages. In Argentinian Spanish, *tipo* as a quotative marker is often used for 'stereotypical (hypothetical) reported speech' (Mihatsch, 2018:169). Voghera (2013) gives examples of Italian *tipo* marking hypothetical thought or conventionalized text (as in Fig 46) and prefers to refer to this usage as "labelling function" since the prevalent discourse function is not for actual quotation (2013:297).

> *formule* **tipo***: "Spingi che ce l'hai fatta"*
> '*clichés* **like** "Push and you've almost made it"'
> **Fig 46: Reproduced from Voghera (2013:297)**

In this thesis a broader definition of the quotative is employed, accommodating imaginary, quasi- and actual quotation. The development of new quotatives is a perfect illustration of one of the hallmarks of grammaticalization processes, subjectification, since the speaker often utilizes **mimesis** to divulge their attitude towards the person they are quoting. Buchstaller (2014:409) explains how 'quotes allow the reporter to superimpose evaluative stances', and also to perform the quote as if acting.

In Russian, the quotative function of *tipa* rather than simply marking reported speech conveys the added semantic nuance of distancing the speaker from the epistemic certainty of the utterance (thereby combining an approximative element).

7.2.9 Functions of type cognates: focus marker, filler and other discourse pragmatic functions

Discourse marker and Focus marker

The term 'discourse marker' has been widely employed in studies in the past 30 years and depending on the individual scholar and their approach subsumes a whole variety of pragmatic uses. These can be subdivided into more specific categories separating out discourse management uses from fillers, clause-final tags, focus markers and connectors. Another broad term 'pragmatic marker' can also accommodate some, if not all, of these uses, but is less preferred here because the approach in this thesis is to view all communication as pragmatic. 'Discourse marker' or 'discourse pragmatic marker' will therefore be used as the broad term to subsume the (functions of) lexemes which are not truth-propositional but are employed by a speaker to convey subjective attitudes and structure discourse.

There is merit in defining the various subcategories of discourse markers, and indeed whole studies have been conducted on individual functions. Here the emphasis will be on those functions which are most pertinent to *type* and *aisa*: focus marker and filler.

Focus marker

Focus markers are defined by Ekberg, Opsahl and Wiese (2015:99) as 'an expression whose referent is focused at the level of information structure… [and] do not contribute any additional semantic content'. Focus markers highlight rhematic content, i.e. the pragmatically most significant information in a sentence.

Examples from studies of *type* include Fig 47 from Argentinian Spanish and Fig 48 from Italian. In the Argentinian Spanish example, *tipo* focuses the hearer's attention on the salient part of the utterance, the adjective *linda* 'pretty'.

> *porque es **tipo** asi re linda*
> 'because she is **like so** very pretty'
> **Fig 47: Copied from Mihatsch (2018:169, example 28)**

The Italian example in Fig 48 illustrates how *tipo che* lit. 'type that', a compound formed from the *type* cognate and a complementizer, functions as a turn-initial focuser (Voghera, forthcoming).

> ***Tipo che*** *se tu fossi cibo io sarei obesa*. (Twitter)
> **Like** if you were food, I would be obese'
> **Fig 48: Copied from Voghera (forthcoming, example 65)**

Note that Fig 47 also includes a combination of *tipo* with (grammaticalized) similative demonstrative *asi* 'like this', here fulfilling an intensifying function. Section 7.3 in this chapter presents research on the grammaticalization of similative demonstratives and section 7.4 explores the combining of these two classes of lexemes with varied outcomes.

Filler

Gap fillers (henceforth fillers) are words which are used instead of silence to fill a gap between utterances or parts of an utterance: 'non-silence devices that can be deployed after the current word has been brought to completion to delay the next word due' (Fox, 2010:2). Fillers are generally understood to be semantically empty although pragmatically they fulfil the function of holding the floor rather than yielding one's turn to another speaker. The use of fillers grants the speaker more time to retrieve the word they were searching for or regain their train of thought.

Russian *tipa* has become so salient as a filler that its usage was ridiculed on the Russian website Absurdopedia (Kolyaseva and Davidse, 2018:217).[55] The example in Fig 49 from Russian *tipa* cannot be a quotative usage since quotatives always immediately precede the quoted text; here, *tipa* is unequivocally a gap filler.

> [muž, 22] *Nu, ona* **tipa** *skazala, čto ja s nej ne xoču bol'še vstrečat'sja iz čuvstva mesti.* [m, 22] 'Well, she **like**, said that I don't want to date her out of revenge.'
> **Fig 49: Copied from Kolyaseva and Davidse (2018:217, example 69)**

Kisiel and Kolyaseva (forthcoming) remark that the use of *tipa togo* ('something like that') as a discourse marker or filler has become a 'verbal tic' for many speakers. Although it also functions as a hedge, its use next to a hesitation marker in Fig 50 supports the filler interpretation.

> | *Mne* | *ešče* | *xotelos',* | *èto,* | ***tipa togo,*** | *zakazat'…* |
> |-------|--------|-------------|-------|------------------|-------------|
> | I-DAT | more | want-PST.N.SG | HESITATION MARKER | **TIPA TOGO** | order-INF |
>
> 'I also wanted, err, **sort of**, to order…'
> **Fig 50: Copied from Kisiel and Kolyaseva (forthcoming, example 65, bold mine)**

'Filler' appears at the far end of the table in Table 1 because the filler is arguably the most semantically bleached and most highly grammaticalized of all uses of *type* cognates. Fillers are not syntactically restricted and function externally of the sentence structure. In English,

---

[55] See https://wikiindex.org/Absurdopedia.net for more details and a link to the website.

type noun phrases *kind of* (*kinda*) and *sort of* (*sorta*) and similative *like* fulfil many of the functions listed in Table 1, including as fillers.

Other discourse pragmatic functions

There are a host of other discourse pragmatic functions attested by grammaticalized *type* cross-linguistically. There is not space to do justice to all, but an example from Russian will suffice as an illustration.

Russian quotative *tipa* has further developed as a falsity marker, indicating that the statement being uttered is not true (Fig 51).

> *no ja xot' i molču i voobšče **tipa uperlas'** v svoj telefon*
> 'but even though I am keeping quiet, and am **"busy"** with my phone…'
> **Fig 51: Excerpt from Kisiel and Kolyaseva (forthcoming, example 60)**

As a falsity marker, *tipa* thus fulfils a similar function to 'air quotes' (represented by inverted commas in written text).[56] Also termed as a 'quasi' use, the falsity marker is not limited to its use alongside quotations, as is evident from the example in Fig 51 in which the speaker, a mother secretly observing her child playing with another child, deliberately says something that is not true whilst highlighting the fact that she knows it is not true. Ukrainian *typu* also fulfils the same function (Kisiel and Kolyaseva, forthcoming).

---

[56] 'Air quotes' refers to the gesture performed with index and middle fingers of each hand whilst speaking the word or phrase. The gesture provides a physical representation of inverted commas used in written text and indicates that the word or phrase uttered is intended to be understood as false, i.e. meaning the opposite of what is spoken.

7.3 Grammaticalization of similative demonstratives cross-linguistically

7.3.1 The similative demonstrative: a special kind of demonstrative

Diessel's (1999) influential monograph on demonstratives draws from 85 languages from multiple language families to categorise demonstrative determiners cross-linguistically along common semantic, syntactic and pragmatic lines. Diessel describes four main deictic uses of demonstratives, the first of which is understood to be the original and basic function: **exophoric** reference to the physical world, often accompanied by a gesture pointing to the relevant object. **Endophoric** (anaphoric, discourse deictic and recognitional) uses are all assumed to have developed from this primary function and are all used with discourse rather than physical referents: **anaphoric** reference is to a prior mentioned noun; in the **discourse deictic** function, demonstratives highlight an aspect of a discourse proposition (which could be anaphoric or cataphoric, appearing after the demonstrative); **recognitional** demonstratives always co-occur with a noun and refer to an entity known by both speaker and hearer, e.g. 'that game Dad bought you'. As is common cross-linguistically, Hindi has deictically contrastive demonstratives for proximal and distal referents, यह *yah* 'this' and वह *vah* 'that', which are inflected to the 'oblique' form if followed by a postposition, becoming इस *is* 'this' and उस *us* 'that'. In addition to simple demonstratives many languages (including Hindi) also exhibit prolific use of another kind of demonstrative determiner, the similative demonstrative.

While simple demonstratives point to a <u>specific</u> referent, e.g. '**that** car', '**this** jacket', similative demonstratives point to a <u>similar</u> referent, combining the semantic-pragmatic function of similative comparison with the deictic function of a demonstrative determiner, as in German *so* 'a car **like that**' or '**this kind of** jacket'. English *such* is the equivalent of German *so* but its usage as a phoric modifier has become antiquated and it has all but lost its exophoric usage (König and Umbach, 2018:289), having been replaced by complex demonstratives such as 'like this/that'. *Such* still appears with extended uses as an intensifier or booster (*Such a beautiful view!*), an exemplifier (*such as…*) and a general extender (*and suchlike*).[57]

---

[57] Note that German *solch* is a slightly outdated alternative to *so* in the syntactic environment preceding the noun, as in this example from König and Umbach (2018:288): *So ein Auto/Ein solches Auto* (+ pointing gesture) *möchte ich.* 'I would like to have a car like that./That's the kind of car I would like to have.' *Solch* has not grammaticalized to the extent of *so* and therefore will not feature in the rest of the chapter.

As with demonstratives, similative demonstratives are also distinguished along the proximity dimension in some languages, with a two- or three-way distinction. The example from Odia in Fig 52 illustrates a two-way distinction between proximal and distal similative demonstratives.

Odia	mora	**emiṯɪ**	ga:ɖi-ṭe	nuha~,	**semiṯɪ**	baɖa
	ga:ɖi-ṭe ɖaraka:ra					
gls	my	**SD.ADN.PROX**	vehicle-CL	be.NEG	**SD.ADN.DIST**	big
	vehicle-CL	required				
eng	I need not a car **like this** but a van **like that**.					

Fig 52: Copied from Sahoo and van der Auwera (2019:22, Fig 9, my translation)[58]

Following an investigation of the semantic properties of similative demonstratives, Umbach and Gust (2014) conclude that 'using similarity demonstratives generates similarity classes… which can be considered as ad-hoc kinds' (2014:75). Thus in the phrase 'I want such a car', an ad hoc kind (or category) is evoked comprising cars similar to the one being pointed at. This is the case regardless of whether the 'pointing' is literal or endophoric. The features or properties highlighted through the similative comparison can be different depending on the context. For example, in Fig 54, which we understand as exophoric, the level of detail of the comparison could be limited to the size of vehicle. The comparison may be between different kinds of people, focusing on features or qualities they might possess. In this case, a similative demonstrative could produce a similarity reading or a categorizer (kind) reading: 'a colleague like that' or 'that kind of colleague'. On the other hand, the sentence from Telugu (Fig 53) involves a comparison between two phones and is more likely to trigger a categorizer reading as it references established categories of phones, specifically of the iPhone.

Telugu	*iṭuvanṭi*	*pho:nu ka:ɖu,*	*na:-ku*	*aṭuvanṭi*	*i-phone ka:va:li*
gls	SD.ADN.PROX	phone not	me-to	SD.ADN.DIST	i-phone required
eng	Not **this type of** phone, I want **that type of** i-phone.				

Fig 53: Copied from Sahoo and van der Auwera (2019:22, Fig 5b)

These examples all highlight the cognitive basis for the inherent connection between type nouns and similative demonstratives. The latent similarity reading within every type noun context is indeed credited with its development as a similative (Voghera, 2013).

Similative demonstratives - the content dimension

The category 'similative demonstrative' has only recently been proposed, in Sahoo and van der Auwera (2019) and van der Auwera and Sahoo (2020), using data from languages as

---

[58] Here and elsewhere, SD refers to a similative demonstrative.

diverse as Sanskrit and Kannada, Latin, French and English to form their argument. Other terms used include 'MQD demonstratives' (König, 2017; König and Umbach, 2018) and 'similarity demonstratives' (Umbach and Gust, 2014; 2021). König's nomenclature is based on his seminal work dividing similative demonstratives in terms of the so-called 'content dimension' denoting the area of reference of the lexeme, either M(anner), Q(uality) or D(egree).

Fig 54 illustrates a similative demonstrative of manner, describing the manner in which an action is performed. The exophoric usage is accompanied by a gesture.

> *Karl böyle koş-uyor.*
> Karl like.this run-PRES.3SG
> 'Karl runs like this (+ gesture)'
> **Fig 54: König and Umbach (2018:373, example 20a)**

The sentences in Figs 52 and 53 are examples of quality demonstratives. Similative demonstratives of quality draw a comparison based on a specific quality of the object being referred to. In Fig 53, the similative demonstrative refers unambiguously to a similar model of iPhone, whereas in Fig 52, it may be a similar appearance or style of vehicle. The specific feature of comparison is inherently vague, relying on the hearer to draw the comparison on the basis of visual input, or a discourse description.

An example of such an endophoric use of Brazilian Portuguese *assim* is in Fig 55, from Lopes and Carapinha's (2004) comprehensive study based on data from a synchronic corpus (written and oral). The deictic referent is anaphoric, referring to milk described as 'foam[y]' and 'sweet' in the first part of the sentence.

> *Tinha espuma, era doce, nunca mais voltei a beber leite **assim*** [L0426P00010X]
> It had foam, it was sweet, I never drank milk [**like that / of that kind**] again.
> **Fig 55: Copied from Lopes and Carapinha (2004:3, Fig 4)**

The authors describe *assim* as an adverb of manner (which would render a prototypical meaning of 'like this; in this way'), yet acknowledge that there are frequent uses of *assim* in the corpus that are not recognized by grammars and fulfil more of a predicative function. One such example is shown in Fig 56 and it clearly shows *assim* functioning as a similative demonstrative of kind and not of manner. The authors astutely note: 'It seems to us that [*assim*] is the only adverb of the language capable of assuming this syntactic behaviour.' Perhaps the reason *assim* is able to fulfil this syntactic role is because in this context, as part of an NP, it is not in fact an adverb, but rather a similative demonstrative.

> *Os meninos pequenos às vezes são **assim.*** [L0112P0021X]
> Little boys are **like that** sometimes.
> **Fig 56: Copied from Lopes and Carapinha (2004:3, Figure 3)**

It may be that the manner adverb was the earliest usage and the kind reading developed subsequently. Or it could be that the authors simply did not distinguish the two since the same word *assim* is used for both.

An example of a degree demonstrative (polysemic German *so*) is in Fig 57.

> *Karl ist **so** (+ gesture) **groß.***
> 'Charles is **this tall / as tall as this**.'
> **Fig 57: From König and Umbach (2018:374, example 22a)**

As in Brazilian Portuguese, languages do not always differentiate between the different members of what König and Umbach (2018) refer to as the 'content dimension'. In terms of a two-way distinction, languages such as Spanish and Portuguese conflate manner and quality and have a distinct lexeme for degree, whereas Polish conflates manner and degree and has a distinct lexeme for quality. Armenian has a three-way distinction in both the proximity and the content dimensions, as does Japanese (see König and Umbach (2018) for details).

Sino-Tibetan language Yakkha makes a three-way distinction in terms of proximal, distal and anaphoric. The anaphoric form of the demonstrative means something akin to 'like mentioned before' and there are forms for time, location and size 'as much as mentioned before' as well.[59]

7.3.2 Overview of the semantic-pragmatic functions of similative demonstratives

Although similative demonstratives (*such* and *so*) now have significantly reduced functions in English, they are frequently used in other languages with both the original meaning and highly grammaticalized meanings. Research on the grammaticalization of similative demonstratives is only available for some lexemes and languages, yet those studies which do exist reveal a remarkable consistency with the kind and variety of new semantic-pragmatic functions developed by *type* and its cognates (and indeed other type nouns).

Table 2 follows a similar method and principle to that of Table 1 and depicts the semantic-pragmatic functions of similative demonstratives based on a cross-linguistic survey of research in several languages.[60] As in Table 1, those functions which pertain to Hindi are

---

[59] See website https://languageswithwilf.wordpress.com/2020/08/15/chapter-4-yakkha/
[60] Note that the paper on Estonian *nii* (similar to the paper on German *Typ* versus *Art*) investigates a particular issue and may not document every function of *nii*.

those which have been delineated. The addition of one language from each of the Uralic and Semitic families enables us to expand the cross-linguistic picture presented in the previous section.

Source material for cross-linguistic overview of similative demonstrative functions

Germanic languages

English *such* and *so*: König (2017, 2020); König and Umbach (2018); van der Auwera and Coussé (2016)
German *so*: König (2017, 2020); König and Umbach (2018); Overstreet (2005); Wiese (2011)
Danish *sådan*: Schoning, Helder and Diskin-Holdaway (2021)
Norwegian *sånn*: Ekberg, Opsahl and Wiese (2015); Johannessen (2012); Vindenes (2018)
Swedish *sån*: Ekberg (2010); Ekberg, Opsahl and Wiese (2015); Johannessen (2012)

Romance languages

Brazilian Portuguese *assim*: Bittencourt (1999); Lopes and Carapinha (2004); see also Lopes-Damasio, L.R. (2008)

Semitic languages

Hebrew *kaze*: Maschler (2001)

Uralic languages

Estonian *nii*: Keevallik (2005)

Language	Word	Phoric modifier	Phoric similative (quality)	Phoric similative (manner)	General extender	Intensifier	Exemplifier	Hedge	Quotative	Filler	Focus marker	Other discourse pragmatic
English	*such*	(✓)			✓	✓	✓					✓
	so			(✓)	✓	✓		✓				✓
German	*so*	✓	✓	✓	✓	✓		✓	✓		✓	✓
Danish	*sådan*	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Swedish	*sån*	✓				✓		✓	✓		✓	✓
Norwegian	*sånn*	✓	✓	✓		✓		✓	✓	✓	✓	✓
Brazilian Port.	*assim*	✓	✓	✓			✓	✓	✓	✓	✓	✓
Hebrew	*kaze*	✓		✓	✓			✓	✓		✓	
Estonian	*nii*			✓					✓			✓

Table 2: Cross-linguistic comparison of semantic-pragmatic functions fulfilled by similative demonstratives

It will be clear from the data in Table 2 that many of the extended uses of similative demonstratives are the same as those of type nouns, notably <u>all</u> of the functions on the more grammaticalized right-hand side of the chart: exemplifiers, hedges, quotatives, and different kinds of discourse pragmatic markers including fillers and focus markers. Many of these extended uses develop from a similative reading of the type noun, showing a strong and obvious cognitive connection to the similative demonstrative and vice versa, as König and Umbach explain (2018:304):

> 'The semantic analysis of manner, quality and degree demonstratives starts out from the idea that they express similarity… Since similarity is, from a cognitive point of view, basic in classification processes, sets of similar items – "similarity classes" – appear closely related to kinds. It will in fact be argued that **the similarity classes created by manner and quality demonstratives are ad-hoc created kinds**.'

In what follows, examples from the full-length studies analysed help to illustrate how each function is fulfilled by similative demonstratives. Commentary is kept to a minimum as most of the categories have already been introduced.

7.3.3 Functions of similative demonstratives

Phoric modifier

In section 7.2.3, the phoric modifier construction was defined as a noun modifier which focused attention on the object in question, the instance, rather than (solely) on the category. When the similative demonstrative is employed as a phoric modifier, the similative sense is more accessible than with type nouns, as the dual translation of Fig 58 indicates.

> (speaker pointing to a mug on the table): ***So eine Tasse** hat Anna auch.*
> 'Anna has **such a mug / a mug like this**, too.'
> **Fig 58: Copied from König and Umbach (2018:13, Example 31b)**

Fig 58 is an example of exophoric deictic reference. Diessel (2006:475) comments on the transformation of demonstratives from providing exophoric to endophoric deictic reference: 'the deictic centre is shifted from the physical world, i.e. the speaker's location at the time of the utterance, to a particular point in the unfolding discourse.' Similative demonstratives achieve this shift as phoric modifiers by focusing attention on and drawing similarity with an element in the discourse, such as in the example in Fig 59 from Hebrew in which similative demonstrative *kazé* combines with a vague noun *davàr*, clearly referring anaphorically to something just said (see also Andersen, 2010, on vague nouns).

> *'az 'al **kazé davàr**, nifradim*?
> So because of **a thing like this** one breaks up?
> **Fig 59: Copied from Maschler (2001:301, excerpt 4)**

Phoric similative

Without a specific referent for the demonstrative component of the similative demonstrative in the immediate discourse context, the effect created by the phoric similative in Fig 60 is a vague and general reference to the preceding discussion (anaphoric), or to shared knowledge held by speaker and hearer (recognitional).

> *De er **sånne***
> They are **like that**
> **Fig 60: Copied from Odden (2019:161, example 66b; translation mine)**

Intensifier

In English, the intensifying use of *such* (as in *It's such a nice day!*) is now much more frequent than its function as a phoric modifier. As an intensifier, the similative demonstrative increases the degree of the item under discussion; in Fig 61, this is the helpfulness of Mrs Jones, and in Fig 62, the degree of luck.

> *Mrs. Jones is always so/sooo helpful.*
> **Fig 61: Copied from König (2020:34, example 14f)**

> *sån tur jag hade avstängt*
> 'such luck I had turned it off' (the mobile phone)
> **Fig 62: Copied from Ekberg (2010:54, example 5)**

Hedge

Wiese (2011) argues that the hedging use in bridging contexts like Fig 63 has led to German *so* being further interpreted as a focus marker.

> *Diese Literatur ist nicht **so** ganz leicht lesbar*
> This literature is not, **like**, entirely easy to read.
> **Fig 63: Copied from Wiese (2011:1023, example 25a)**

Quotative

With type nouns, it was the similarity and approximative readings that gave rise to their usage as quotatives, with reported speech understood to be similar to the original but not exact. It is logical that the similative aspect of similative demonstratives would cognitively give rise to the same innovations. Buchstaller and van Alphen's (2012) volume draws together research on these so-called new quotatives cross-linguistically and illustrates how 'similarity/approximation' and 'demonstrative deictics' are the most prolific semantic sources of new quotatives. The productive pathway from similative to quotative is included

in the World Lexicon of Grammaticalization (Kuteva et al., 2019), particularly highlighting the manner similative.

While proximal and distal similative demonstratives are both used in endophoric constructions, the quotative tends to develop from cataphoric uses of the proximal similative demonstrative (König, 2017; 2020), as illustrated in Fig 64. It is only a short gap from the (original) manner demonstrative meaning of Estonian *nii* to a quotative marker for what the author terms 'vocal demonstrations'. The imitative aspect of the quotative is foregrounded as the vocalisation that is produced is understood to be in a 'similar manner' to the original.

> *jah teate küll **teeb nii** < ruii ruii rui rui [rui ruii] >*
> 'yeah, you know it **does like this**, "oink oink oink"' (punctuation mine)
> **Fig 64: Copied from Keevalik (2005:116)**

Note that *nii* 'like this' always appears before the imitative sounds which follow cataphorically and is similar to English quotative *like* in that it does not appear on its own without another verb. It usually appears in conjunction with a light verb such as 'do', as in the pig imitation example.

Focus marker

In Fig 65, the Swedish similative demonstrative *sån* appears immediately before *busfrö* 'little devil' and highlights its rhematic importance.

> *hon ville inte ha mej där bak asså jag var **sån** busfrö*
> 'she didn't want me to sit in the back [of the room] because I was, **like**, a little devil'
> **Fig 65: Copied from Ekberg (2010:56, example 9, translation mine)**

Discourse-structuring functions

Similative demonstratives have a broad application as discourse-structuring devices, including their development into sentence-initial adverbial discourse markers and also formulaic fixed expressions which mark the beginning or end of a turn (see König, 2020, on formulaic expressions in English and German).

Fig 66 illustrates the discourse management function of Brazilian Portuguese compound *tipo assim.* Use of the compound is polite and possibly has a mitigating function; it seems the speaker wants to avoid appearing argumentative or saying anything against the other person's preference.

<blockquote>
A: *agora eu tô na do forró …*
Now I'm into *forro*.

B: **tipo assim** *… forró não faz meu gênero … meu negócio é rock...*
Well … *forro* isn't my kind of thing … my thing is rock…
</blockquote>

Fig 66: Reproduced from Bittencourt (1999:50, example 22a)

Other discourse uses

As with *type* cognates, the extended discourse functions of similative demonstratives are proliferous. To name but one example, English and German *so* have developed adverbial connectives which fulfil a variety of discourse functions, including resultative (as in Fig 67).

> I would like to wait, **so** that I can get home dry.
> **Fig 67: Copied from König and Umbach (2018:9, example 26d)**

7.4   Interaction between type nouns and similative demonstratives

Having surveyed the landscape of *type* grammaticalization cross-linguistically and ascertained that there is substantial overlap in the functions developed by grammaticalized similative demonstratives this section explores the interrelationship between them in the broader context of the interaction between grammaticalized type nouns, similatives and similative demonstratives in general. The cognitive and syntactic bases for semantic innovation within individual type nouns and similative demonstratives have already been presented. Factors influencing the development of these lexemes are discussed, including the principle of attraction which operates between 'competing' lexemes (De Smet et al., 2018) and the history of a lexeme, including its etymology and the constructional family to which it belongs. Specialization is one outcome of competition between lexemes within a functional context, and examples from several languages are presented which illustrate how type nouns have specialized vis-à-vis alternative lexemes.

7.4.1   Clusters and compounds

A feature of grammaticalized uses of type nouns and similative demonstratives is that different lexemes with the same or similar function tend to cluster together. Clustering may therefore be an outcome of attraction between similar lexemes or it may (also) be the semantic-pragmatic context which attracts multiple lexemes to itself: attraction between 'strategies'. In the example in Fig 68, type noun *genre* acts as a focus marker, focusing *zonard* 'dropout', whereas *style* fulfils a hedging function (Fleischman and Yaguello, 2004:138).

> *C'est un mec GENRE STYLE zonard.*
> He's KINDA LIKE a dropout.
> **Fig 68: Copied from Fleischman and Yaguello (2004:138, example 47)**

However, the authors acknowledge that there is 'functional overlap' between *genre* and *style*. This corresponds with the conclusion of other scholars that innovative uses of type nouns can fulfil more than one function simultaneously and it is not always possible to distinguish between them.

The quotative context also seems to attract clusters of quotatives with other quotatives and with hedges. This may be due to the origin of new quotatives in similative and approximative lexemes, which indicate that the quote is an approximate rather than an exact rendition. Innovative uses of type nouns cluster with other innovative discourse-pragmatic strategies such as new quotatives. Mihatsch also notes that '[a]pproximation markers are often stacked' and proposes reasons that include word retrieval and 'flagging problems in the formulation process' (Mihatsch, 2020: 145). The inherent attraction between lexemes with a similar or complimentary pragmatic function could equally provide an explanation.

(15) *MAR: *no // en el comedor / lo tiene [///] como no lo tenían puesto … *VIS: hhh y entonces luego / hhh / o sea / es [/] es **como** / <|tipo|> cuartito / con el sofá / la tele / **y eso** // *(C-ORAL-ROM)*
'MAR: no // in the dining room / he has it [///] since they didn't have it on … *VIS: hhh and then / hhh / that is / it's [/] it's **like** / <|**kind of**|> a little room / with the sofa / the TV / **and that** // *(C-ORAL-ROM)
Fig 69: Copied from Mihatsch (2020: 153, Figure 38)

The fact that it is the non-nominal uses of type nouns (and non-deictic uses of similative demonstratives) which tend to cluster (as in Fig 69) can perhaps be explained by the relative syntactic flexibility of adverbs and discourse particles which are often additional to the core structure of a sentence and can be stacked; this is in contrast to prepositions, without which the sentence would often be ungrammatical and which almost never appear in clusters.

Clusters of type nouns and similative demonstratives

A phenomenon affecting non-nominal uses of *type* is that of clustering, wherein the discourse-related function of the type noun (including *type*) is reinforced by its cooccurrence with other words of a similar function, such as the clustering of two or more hedges or a hedge and a booster.

Marques' (forthcoming) study of combinations comprising *tipo* and other discourse markers in European Portuguese uncovers the following three main syntactic structures:

1. Discourse marker + *Tipo*
2. *Tipo* + Discourse marker
3. Discourse marker + *Tipo* + Discourse marker

Combinations with *tipo* and similative demonstrative *assim* 'like this' are the most frequent (forthcoming). In Brazilian Portuguese, this combining of *tipo* and *assim* is shown to be the result of a process by which *tipo* and *assim* gradually become part of a syntactic compound (Lopes and Carapinha, 2004). However, Marques (forthcoming) claims that in European Portuguese whilst *tipo* appears in a cluster it never forms a compound, i.e. never forms a syntactic unit with its counterpart. Fig 70 is an example of a cluster in European Portuguese; *tipo* functions as a(n approximative) quotative, and *assim* acts as a hedge on the whole phrase.

> *as instalações são porreiras / mas é **assim** / eh / **tipo** que / queres alguma coisa / tens que ir lá*
> 'the facilities are cool / but it's **sort of** / eh / **like** / [if] you want something / you have to go there'
> **Fig 70: Copied from Marques (forthcoming, Fig 36)**

The example in Fig 71 shows Swedish *typ* clustering with Swedish *sån*. In this utterance, *typ* can be interpreted as a discourse marker which accentuates the phrase following it. The quantity *två gånger* 'twice' is itself brought into focus by *sån*, which Ekberg, Opsahl and Wiese (2015:58) describe variously as a discourse marker, a 'focusing discourse marker' and a 'focalizer'.[61]

> *byta **typ sån** två gånger*
> exchange **type such** two times
> 'exchange **like** two times'
> **Fig 71: Ekberg, Opsahl and Wiese (2015:56, example 10)**

Both discourse and focus marking, and indeed hedging and exemplification, are fulfilled by the English word *like*, hence the use of a single word in the translation. This highlights the fact that lexemes can fulfil multiple functions simultaneously, a feature which we also observe with *type* cognates.

In Czech *typ* commonly clusters with the similative demonstrative *takový* 'such; like this', as in Fig 72, producing an ad hoc category (Janebová et al., forthcoming).

---

[61] Note that *sån* appears in the singular even though *två gånger* is plural (Ekberg, Opsahl and Wiese, 2015). This highlights its function and status as a discourse marker and not a similative demonstrative in this context.

takový	*rozervanec*	***takový***	***ten***	***typ***	*Karel*
such	savage	**such**	**that**	**type.NOM.SG**	Karel.NOM.SG

'such a romantic savage, **such as** [the Czech poet] Karel Hynek Mácha'

Fig 72: Janebová et al. (forthcoming, example 13)

These clusters are perhaps a result of similar meanings attracting each other and seem to have become conventionalized. Is the Czech cluster of *takový* with *typ* truly a cluster or rather a compound? Further research is required in order to determine this. One clearly identified compound is Brazilian Portuguese *tipo assim*, which will be discussed in the next section.

Fixed combinations of a type noun and a similative demonstrative

Brazilian Portuguese exhibits not only grammaticalization of taxonomic noun *tipo* 'type' and similative demonstrative *assim* 'like this' (as illustrated earlier in the chapter), but also the use of both lexemes together in a compound *tipo assim*, with a variety of similar pragmatic functions. Bittencourt's (1999) analysis of the grammaticalized uses of Portuguese categorizer noun *tipo assim* is based on an informally gathered corpus of spontaneous speech from (mainly) adolescents and youth, with some examples collected from adults and from TV programmes. The author identifies several functions of the prolific compound: hedges (of both quality and quantity), exemplification, complementizer, quotative, and various uses in discourse management, to introduce a new topic as well as close a topic, to express agreement as well as disagreement, and to indicate hesitation, used as a filler.

Bittencourt illustrates the possible process by which the two lexemes could have combined to form a fixed expression through repeatedly appearing in proximity, first in the same phrase, then in the same phonological unit (exemplified in Fig 73) and finally juxtaposed.

*eu mandei ele fazê uma lista de palavras ... **tipo** essas **assim** que o pessoal da informática tá usano*

I made him make a list of words … **like** those **(kinds)** which computer people are using

Fig 73: Reproduced from Bittencourt (1999:45, example 8a)

Ultimately, *tipo assim* is observed functioning as a single unit, such as in the approximative use in Fig 74 where *tipo assim* hedges the locational phrase *lá fora* 'outside'.

*cê me Leva e fica me esperano **tipo assim** lá FOra ...*

'Take me [in the car] and wait for me, **like**, outside.'

Fig 74: Reproduced from Bittencourt (1999:45, example 9b)

What is especially pertinent is that all the same functions can be fulfilled by *tipo* on its own (Bittencourt, 1999) or by *assim* on its own (Lopes and Carapinha, 2004), as the first two sections of this chapter indicate. A recent study of a Brazilian Portuguese youth dialect from

Rio de Janeiro (Thompson, 2019) also listed *tipo assim* as synonymous with *tipo*. As of now, no study has analysed the specific lexical and pragmatic contexts in which either of *tipo* or *tipo assim* is preferred above the other. This would be a worthy exercise likely to highlight subtle differences in lexical semantic meaning and/or pragmatic usage between the two.

Brazilian Portuguese offer an example of a (grammaticalized) type noun combining with a (grammaticalized) similative demonstrative to form a new linguistic item. The collocation *tipo assim* is not attested in European Portuguese, nor is its cognate *tipo asi* attested in Peninsular Spanish, although it has been attested in Argentinian Spanish, possibly through the influence of Brazilian soap operas, which are very popular in Argentina (Mihatsch, 2018).

In Hebrew, type noun *min* 'sort, type' combines with similative demonstrative *kaze* 'like this' to form a fixed expression *kaze min X* which translates as 'this sort of X' and has a hedging function (Maschler, 2001:307). This is illustrated in Fig 75; the boy is described as 'this sort of professor', indicating that he is studious.

> *'az hayéled haze, … 'eh **kaze min profésor**,*
> So this boy, [is] uh **this sort of professor**,
> **Fig 75: Reproduced from Maschler (2001:307, Excerpt 10)**

The strong interrelationship between *tipo* and *assim* is evidence of the cognitive connection between type nouns and similative demonstratives, a theme which will recur throughout this thesis as the relationship between loanword *type* and native Hindi lexeme *aisa* is explored. In addition to shedding light on similatives (and similative demonstratives) of quality and their role in grammaticalization processes, this research will contribute to knowledge of the recently defined area of similative demonstratives and their grammaticalization. In a rare phenomenon documented elsewhere only in Norwegian, this study shows how a *type* construction with the demonstrative becomes a complex similative demonstrative, a grammaticalization process not hitherto evidenced in other languages. The cause for this development is proposed to be attraction between the type noun and the similative demonstrative coexisting in the lexicon, which utilizes the mechanism of analogical replication. The next section introduces the theory behind this proposition.

7.4.2   Competition and attraction

We have observed interaction in the form of clustering and compounding of type nouns with similative demonstratives. Another significant form of interaction between the two groups of lexemes can be explained using the principle of attraction, an alternative to competition theory proposed by De Smet et al. (2018) and described in Chapter 4.6. Among the literature

surveyed, there is one language in particular in which the principle of attraction as described by the De Smet et al. (2018) seems particularly well illustrated: Norwegian.

Odden (2019) shows how Norwegian *type* when combined with a demonstrative *den type(n)* 'this/that type' is interchangeable with similative demonstrative *sånn* in its adnominal position. Odden (2019) argues that *den type(n)* has been reinterpreted as a lexical chunk synonymous with *sånn,* following the pattern of another grammaticalized type noun *den slags* 'that kind' which is even sometimes written together as *denslags*. Both fossilized chunks *den type(n)* and *den slags* can therefore be considered complex similative demonstratives, a development which becomes apparent in examples like Fig 76, in which *den type* functions as a phoric similative, without reference to a category noun (type of what?).

> *Jeg vet det finnes **den typen**. De irriterer meg grenseløst.*
> I know you get **that type**. They annoy me limitlessly.
> **Fig 76: Copied from Odden (2019:180, example 93, translation mine)**

Having been thus reinterpreted as a chunk (or compound) rather than a composite phrase, the similarity between *den type(n)* and *sånn* has led to further analogical imitation in which *type*, like *sånn,* fulfils a function as a general extender: 'Here, we may have a kind of functional analogy in which [*den slags (NOUN)*] and [*den type(n) (NOUN)*] are **widening their semantic-pragmatic scope in line with the functions of SÅNN**, and developing further polysemy.' (2019:286, **bold** mine). An example of this prolific innovative usage is shown in Fig 77 which also combines with hedges *sånn* 'like this/that' and *litt* 'a bit'.

> *bor ganske litt langt fra hverandre og i hver sin enebolig **og litt sånn den type***
> live a bit far from each other and in separate houses **and a bit kinda like that** (lit. a little like this/that that type)
> **Fig 77: Copied from Odden (2019:177, example 87, my translation)**

Fig 77 shows clear evidence of decategorialization because *type* is missing its definite suffix *-n*. Grammaticalization of *den type(n)* is further evident by the way in which the determiner agrees with the category noun and not *type* in some sentences.

Finally, *sånn* and *type* frequently combine in the prenominal context where *sånn* alone would suffice (Fig 78). Here the addition of *type* can be understood to add an approximative element to the phrase (Odden, 2019).

> *eller vi vi bruker nok veldig få sånne egne ord # altså **sånn type dialektord***
> 'or we use very few particular words # like **such kind of dialect words**'
> **Fig 78: Copied from Odden (2019:303, example 139a, my translation)**

7.4.3   Functional specialization

Similar to complementary distribution in phonology, where different phonemes appear in different phonological contexts (usage of one thereby 'complementing' the other), we find that lexemes which are ostensibly synonymous can in fact specialize, appearing in different semantic or syntactic contexts, or both. This often entails a subtle difference in semantic meaning. The reasons behind the functional differentiation of two similar lexemes are not always proposed by scholars, but can be understood to be a result of restrictions and influences from the history of a lexeme, either its individual etymology or the constructional family to which it belongs.

One such example concerns German taxonomic nouns *Art* 'kind' and *Typ* 'type' which either appear in different contexts or communicate a different meaning in the same context. As Umbach (forthcoming) explains:

> '*Art* and *Typ* denote two distinct means of classification: classification with *Art* makes use of essential properties shared by the instances of a kind while classification with *Typ* makes use of models or prototypes shaping the tokens of a type.'

This distinction may have existed from the point of borrowing of *Typ* (which came later than *Art*), or the lexical specialization may have developed over time. Regardless of the timeline, it is clear that the appearance of *Typ* led to a reorganization of the semantic field. The concept of competition is pertinent here as new lexemes or forms compete with old forms for usage in the speech patterns of speakers.

This also illustrates that despite the uniformity of *type* cognates fulfilling a categorizer function in all of the languages surveyed (see Table 1), there are subtle differences in the details of how the meaning and usage of categorizer *type* has evolved in each language.

7.5   Conclusions

This chapter has illustrated the striking resemblances between the extended functions of type nouns and those of similative demonstratives, with examples from an increasing number of language families indicating that both groups of lexemes can develop functions as hedges, quotatives, general extenders and discourse markers. The phoric modifier is posited as the pivotal function facilitating this close interrelationship between the two groups of lexemes, since it combines the similative aspect of categorization with a deictic element, the two intrinsic features of similative demonstratives. This is supported by

evidence from Norwegian that the entrenched lexical chunk *den type(n)* 'this/that type' is interchangeable with similative demonstrative *sånn* in most contexts. Section 7.4 showed how the semantic relationship between *type* cognates and similative demonstratives has resulted in combinations in many languages, such as the compounding of *tipo* and *asi/assim* and subsequent grammaticalization of *tipo asi* in Argentinian Spanish and *tipo assim* in Brazilian Portuguese. This highlights the possibility of interaction between the two categories of lexemes as they grammaticalize, an idea which will be explored in this thesis.

The cross-linguistic data yields a comprehensive basis for comparing synchronic uses of loanword *type* as well as those of similative demonstrative *aisa* in Hindi. Since layering is not only common in cases of grammaticalization in general, but more specifically evidenced in all the languages studied here, it is likely that the synchronic data from Pithoragarh will display layering of older and newer forms and functions of *type*. The absence of Indo-Aryan languages from any study on type nouns or type noun grammaticalization means that this study can fill a research gap. Additionally, there has been a dearth in research on pragmatics in Hindi, with no study of hedging devices, general extenders or fillers. Finally, the significance of similative demonstratives in Hindi means that a study of the functions of *type* in comparison with alternative lexemes will yield an opportunity to uncover more cases of attraction between similar and competing lexemes.

8 Synchronic functions of *type* in Hindi

8.1 Introduction to synchronic functions of *type*

8.1.1 Overview of chapter

One of the principal aims of this thesis is to examine the uses of loanword *type* in Hindi within the global context of the grammaticalization of type nouns cross-linguistically; the previous chapter provided the necessary backdrop for this. Another principal aim is to consider the uses of *type* within its local, language-internal context alongside its synonyms in the Hindi language. To this end, the current chapter will introduce the reader to the (previously under-described) strategies available within Hindi for fulfilling each of the semantic-pragmatic functions frequently developed by type nouns, with a focus on those functions performed by loanword *type*. Having laid such a foundation, examples will be presented of *type* fulfilling many of the same semantic-pragmatic functions as the older Hindi lexemes, and with similar syntactic realizations, showing a substantial departure from English uses of *type*. Hindi *type* is found to be multifunctional and highly polysemous, performing the following functions: categorization, manner description, phoric modification, similative comparison (both phoric and non-phoric), hedging, marking quotations, and structuring discourse, including functions as a general extender and a filler. Where relevant, features of the dialect of Hindi spoken in the Kumaon Hills (and exemplified by the Pithoragarh corpus) will be highlighted. In addition to the primary spoken data recorded in Pithoragarh, informal written data from the web will be used to compare with the findings.

The chapter is subdivided into six parts:

Part I Functions of *type*: categorizer and manner descriptor

Part II Functions of *type*: phoric modifier and phoric similative

Part III Functions of *type*: similative

Part IV Functions of *type*: hedge

Part V Incipient functions of *type*: quotative, general extender and filler

Part VI Functions of *type* in web data

Part I to Part V present the functions of *type* and its synonyms as discovered in the spoken Hindi data from Pithoragarh. Clearly, a language with as long and rich a history as Hindi will have many mechanisms for performing communicative functions. Priority will be given to those lexemes which feature in the recordings, with a focus on the uses of similative

demonstrative *aisa* and its related lexemes (*jaisa* and *vaisa*, and their adverbial counterparts *aise, jaise* and *vaise*).

Part VI outlines the uses of *type* identified in the Hindi Web corpus of informal written data from the internet. These are found to be similar as well as different, and the reasons for the differences in particular will be explored in Chapter 10.

The similarities between the synchronic use of *type* and those of native Hindi *aisa* and its related lexemes are striking and will be explored in further chapters. By illustrating how *type* both mirrors the discourse contexts of *aisa* and its related lexemes and appears in identical syntactic environments, the chapter provides a basis for postulating how innovative uses of *type* could have developed; this will be pursued in Chapter 9. Differentiation between the usage of each lexeme in terms of their distribution and collocational preferences will also be briefly explored in Chapter 9.

The following section presents a brief overview of the functions of *type* identified in the spoken data.

8.1.2   Multifunctional polysemous *type*

The data from the Pithoragarh corpus contains a total of 67 tokens of *type*, excluding verbal uses ('to type'). Following a careful analysis of the uses of *type* from both a semantic-pragmatic and a syntactic perspective, 10 distinct semantic-pragmatic functions of *type* have been identified. In three of the functions, *type* can be analysed as a noun, whereas the other seven must be considered functions of non-nominal *type*. These are illustrated in Bar Chart 1.
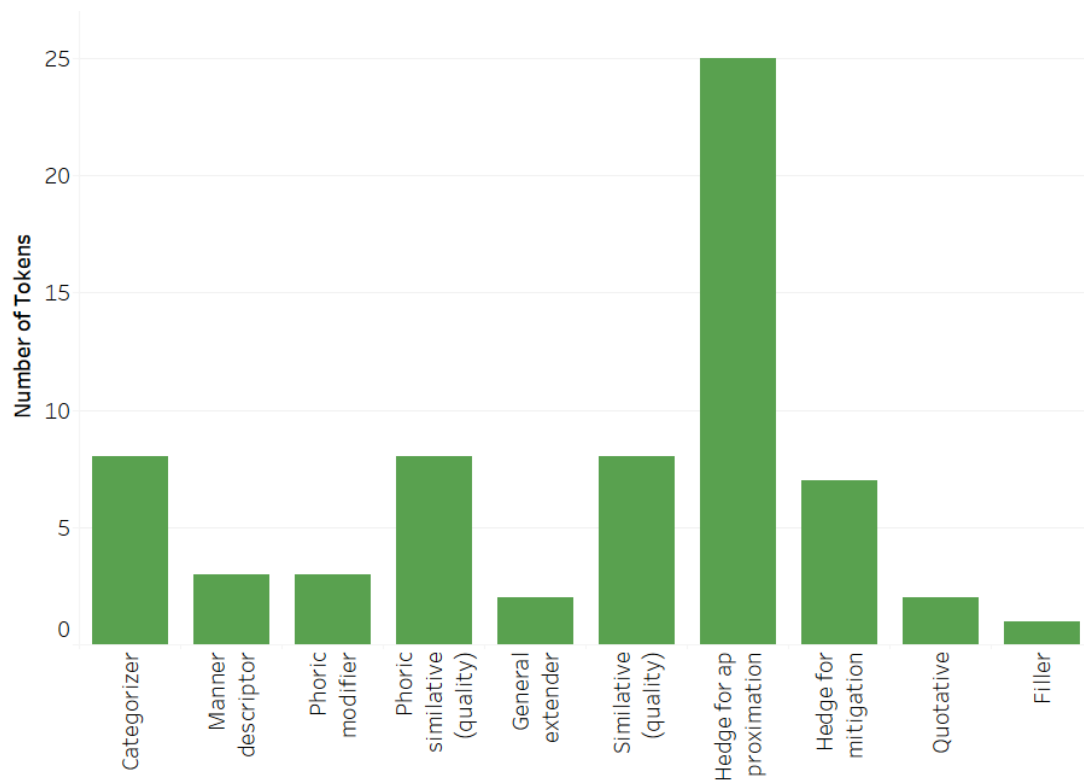
Bar Chart 1: No. of *type* tokens by semantic-pragmatic function

Of the 67 tokens, only eight are taxonomic uses with *type* functioning as a categorizer (e.g. *alag-alag type* 'different types', *kis type ki hindi* 'what kind of Hindi'); the other 59 instances therefore represent examples of semantic innovation or grammaticalization. The semantic-pragmatic functions in Bar Chart 1 follow the order presented in the previous chapter for the cross-linguistic comparison of functions of *type* cognates and are arranged from left to right from most conventional to most innovative, as previously explained (see Chapter 7). This progression coincides with syntactic innovation, moving from nominal to non-nominal uses. The reader will note that these are the same as the semantic-pragmatic functions fulfilled by *type* cognates in other languages across the world as presented in the previous chapter; the same order of functions is followed here, which roughly corresponds to increasing levels of grammaticalization.

The functions of *type* cover all major functions of *type* cognates cross-linguistically with the exception of similatives of manner, exemplifiers and hedges on quantity (rounders). Similatives of quality are attested, and the absence of similatives of manner can be explained by the limited use of *type* as a manner noun. The three tokens of *type* forming a manner adverbial (*vo usi type se bolenge* 'She'll speak in that way') are seen to be negligible and as yet unproductive, although they may represent the start of a future development. Most significantly, demonstrative phrases with *type* are seen to have grammaticalized into a

similative demonstrative which functions as a phoric modifier and a phoric similative in addition to fulfilling incipient functions as a general extender, quotative and filler. The three tokens classified here as examples of *type* as a phoric modifier must therefore be viewed as part of a productive innovation.

Bar Chart 1 highlights the hugely important role played by *type* as a hedge in colloquial Hindi. The use of *type* as a hedge for approximation is by far the most frequent function of *type* in the data and a large section of this chapter (Part IV) will be dedicated to a description and illustration of the ways in which this manifests itself.

That there are fewer tokens on the far right side of the chart may reflect the recency of *type*'s introduction into the language and therefore the limited time in which discourse uses could have developed. Like Hindi, Swedish also attests limited discourse uses of *typ* 'type' beyond hedging and exemplification, with discourse markers only starting to make an appearance in youth language (Rosenkvist and Skärlund, 2013); cross-linguistic tendencies would predict that the quotative and filler uses develop and increase in frequency over time. There are language-internal reasons why *type* may not develop a strong general extender usage; these will be discussed in Part V.

8.2   PART I Functions of *type*: categorizer and manner descriptor

8.2.1   Categorizer nouns in Hindi

In order to investigate the ways in which English taxonomic noun *type* has come to be used in Hindi, this section offers a brief introduction to the Hindi categorizers with which *type* shares a semantic space. Hindi contains several type nouns of varying origins, of which *type* is the newcomer. From Sanskrit origin, and the least developed in terms of innovative extended uses, is प्रकार *prakaar*, which is polysemic in its use as a categorizer noun and as a manner noun (see Fig 79), but less commonly used in informal conversation.

प्रकार *prakaar* [S.] **m.**

2.   kind, sort, type
3.   way, manner, method

Fig 79: Extracts from OHED definition of प्रकार *prakaar*[62]

Fig 80 presents extracts from a conversation which started with Chandrashekar asking his friends which dal (lentil dish) they wanted him to cook for lunch and progressed to a chat about different types of dal.[63] In the middle of a long list, Chandrashekar exclaims, *kai prakaar ki* 'loads of types of [dal]', before continuing with his list of dals and similar dishes.

#	Speaker	Utterance
1	Chandrashekar	*kya khaaoge? masur ki daal?*
	en	What will you eat? *masur* dal?
2	Chandrashekar	*arhar ki? maas ki? bhatt ki daal?*
	en	*arhar* [dal]? *maas* [dal]? Soya bean dal?
3	Bhagwan	**well** *mujhe daalon ke naam bhi jyaada pata nahi hai sach bataaun to*
	en	Well, I don't actually know many dal names of dals, if I'm being honest
	All	--All three speakers list names of dals--
4	Bhagwan	*bhatt ki churkaani*
	en	Soya bean *churkaani*[64]
5	Bhagwan	*bhatt*
	en	Soya bean
6	Chandrashekar	***kai prakaar ki****! bhatt ki daal hoti hai, palyo hoti hai*

---

[62] The Oxford Hindi English Dictionary (OHED) is the main reference dictionary used throughout this thesis.

[63] Note that in Hindi, the same word *daal* is used for the food category of pulses as well as for the dishes made from them. Both meanings also exist in English as borrowings from Hindi and will therefore be used throughout this thesis.

[64] *bhatt ki churkani* is a Kumaoni dish made from black soya beans in a sauce.

	en	**Many (lit. several) types of [dal]!** There's soya bean dal, there's *palyo*[65]
	All	--All three speakers list more names of dals--
7	Chandrashekar	*daal, **tamaam prakaar ki daal** hoti hain*
	en	Dal, there are **innumerable types of dal**! (lit. entire/complete types of dal)
8	Bhagwan	*tu kitne type ki bana leta hai?*
	en	How many types of [dal] do you make?

Fig 80: Bhagwan (M18) and Chandrashekar (M22), convo_e

The glossed version of line #7 is represented in Fig 81, allowing us to observe one of the canonical categorizer structures in Hindi '(ADJ) + TN + GEN + N', highlighted by the blue box. In this sentence, the adjective *tamaam* (here with an expressive meaning of 'innumerable') modifies the type noun (TN) *prakaar*, and is directly followed by the genitive marker *ka* (here *ki* with feminine gender marking) which precedes the category noun (N) 'dal'.

hin	दाल	तमाम	**प्रकार**	की	दाल	होती	हैं
tra	daal	tamaam	**prakaar**	ki	daal	hoti	hain
gls	dal	complete, entire	**types**	GEN.F	dal	be.F	AUX.PL

eng Dal! There are <u>innumerable</u> types of dal!

Fig 81: Chandrashekar (M22), convo_e

The sentence is existential, without a nominal subject, which constitutes one of the main contexts of use of type nouns in Hindi, rendering a binominal type construction similar to the English construction 'type of N'.

Trinominal constructions also exist in Hindi, particularly when giving definitions, as in the formula 'N is a type of N'; the schema can also be written '(SUBJ) [ADJ [*type*]]_{NP} of N2'. Two examples from Hindi Wikipedia illustrate these, with *prakaar* (Fig 82) and with *tarah* (Fig 83).

वाईमनु आम तौर पर **एक प्रकार की पक्षी** की तरह थे[66]
*Waaimanu aam taur par **ek prakaar ki pakshi** ki tarah the*
'[the early penguins called] Waimanu were broadly like **a type of bird**'
Fig 82

---

[65] *Palyo* is a Kumaoni dish made from dried fenugreek leaves in a chickpea flour sauce.

[66] From Hindi Wikipedia entry on 'Penguin' - https://hi.wikipedia.org/wiki/पेंगुइन, accessed on 30th August 2019. Note that *tarah* appears in this sentence in its prepositional form *ki tarah*, yielding a similative sense 'like'.

नकछिकनी - **एक तरह की घास** जिसके सूँघने से छींकें आती हैं[67]

*nakchhikani – **ek tarah ki ghaas** jiske sungne se chhinke aati hain*

nakchhikani (Sneezewort) - **A type of grass** which makes one who smells it sneeze.

Fig 83

Taxonomic sentences occur much less frequently in colloquial discourse than in formal written texts (Mihatsch, 2016). This explains the lack of examples of 'clear categorization' (Vassiliadou et al., forthcoming) in the recordings. When they do occur, they tend to be used for defining unusual or unfamiliar things that cannot simply be named with a common noun.

The more prolific type noun in common parlance, and in many ways the predecessor to *type*, is another Arabic loanword तरह *tarah*, which may have been borrowed into Hindi during the time of the Mughal empire in the fourteenth to sixteenth centuries, along with a vast number of other lexical items still in popular usage. *Tarah* is highly polysemous, and the uses of *tarah* attested in the spoken Hindi from Pithoragarh include the following from the dictionary (OHED) definition of *tarah* partially reproduced in Fig 84.

तरह *tarah* [A.] **nf.**

1. kind, type. इस ~ का, *is ~ ka*, **adj.** 'this sort of'
2. manner, mode, way; *achhi tarah (se)* **adv.** well
5. **ppn.** w. की *ki.* like, as, in the manner (of); *puri ~ se* **adv.** fully, entirely; ~ *tarah* **adv.** badly

---

Fig 84: Extracts from OHED definition of तरह *tarah*[68]

Hindi type nouns do not display all of the properties of a fully lexical noun. Whilst they do show grammatical gender and cause adjectives within their scope to inflect for gender, they do not inflect for number like other Hindi nouns. Instead, they are used (almost) exclusively in the singular, and it is only the adjectives and determiners within their scope that inflect for number. A Google (google.co.in) search on plural phrase "कई तरहों के" *kai tarahon ke* 'several types of' returns only 1,680 hits, whereas using the singular form of the type noun

---

[67] From the Oxford Hindi Dictionary online https://hi.oxforddictionaries.com/definition/नकछिकनी. Accessed on 29th August 2019.

[68] Note that the two Oxford dictionaries used in this thesis are not totally aligned. OEEHD lists only प्रकार *prakaar* and किस्म *kism* as translations of English *type*, although the primary meaning of *tarah* in the OHED is 'kind, type'. Data in the Pithoragarh corpus supports the use of *tarah* as a type noun.

"कई तरह के" *kai tarah ke* returns 2.9m hits (with the same meaning).[69] If the type nouns were ever fully nominal in Hindi, they do not appear to be now. The small percentage of pluralized type nouns which are attested can be explained either as a remnant of the noun's historical grammatical properties, or as hypercorrection on the part of the speaker.

The two most frequent type nouns in the semantic field of categorization apart from loanword *type* are thus both polysemous (*tarah* more so than *prakaar*), although neither has developed uses as extensive as *type*. It must also be mentioned here that Hindi similative demonstrative *aisa* 'of this sort; like this' inherently fulfils a categorizer function indivisible from its deictic similative function in many contexts, as in Fig 85 where, inflected to match the feminine gender of the following noun, *aisa* refers anaphorically to the kind of hospital the three friends are discussing, i.e. an ayurvedic hospital that offers *paanchkaram* treatment.

hin	तो	यहाँ	**ऐसी**	जगह	कहाँ	है
tra	to	yahan	**aisi**	jagah	kahan	hai
gls	so	here	**SD. F**	place	where	COP
eng	But where is there that kind of place here?					

Fig 85: Urmila (F47), convo_g[70]

However, since the similative sense is also indefatigably present ('a place like that'), *aisa* is not considered a true categorizer in this thesis and will be formally introduced in section 8.3 on phoric modifiers.

8.2.2 *Type* as a categorizer

There were eight cases of *type* fulfilling a categorizer role in the recorded data, that is, being used to discuss subcategories of a named category noun. One such example occurs in a conversation about food between three male students. Chandrashekar makes a statement about there being a great many different types of *daal* 'pulses; dal' (see Fig 81 above), after which Bhagwan asks him how many types he cooks (Fig 86). The category noun under discussion, *daal*, was explicitly stated in the previous turn and is therefore not required here.

---

[69] A similar search on 'several kinds of' using plural and singular forms of प्रकार (a masculine noun) revealed similar results.

[70] Note that sentences are translated idiomatically; in Hindi the proximal demonstrative is used for anaphoric reference in discourse, whereas in English the distal ('that') is preferred. Note also that glossed and unglossed examples are employed in this thesis depending on whether the discussion is centred around general context or specific words, how transparent the meaning of the key words is, and whether there are glossed examples in the vicinity which provide transferable information.

Hindi, a pro-drop language, allows the second (category) noun in binominal constructions to be dropped if retrievable from the discourse context.

hin	तू	कितने	टाइप	की	बना	लेता	है
tra	tu	kitne	**type**	ki	bana	leta	hai
gls	you	how many	**types**	GEN.f	make	take.2SG	AUX
eng	How many **types** of [*daal*] do you make?						

Fig 86: Bhagwan (M18), convo_e

The structure of this utterance is the same as that of type nouns *prakaar* and *tarah* described above and should be analysed as '*type* of N' ("TYPE + GEN + N") with the final category noun included. This structure is unequivocal, because the feminine marking of the genitive in Fig 86 agrees with the gender of the dropped category noun.

The following tokens, all from a single utterance by a speaker repeating the phrase *kis type ki hindi* 'what kind of Hindi' (Fig 87), are the only tokens in which the category noun ('Hindi') is explicitly present in the *type* phrase. Even then, it is dropped from the final repetition for dramatic effect.

> *ye hamaare hamaare vajah se iski* [sic] *samajh me aayega ki ye **kis type ki hindi** bolti hai ye **kis type ki hindi** bolti mein **kis type ki***
> She, through us, she (=the researcher) will come to understand **what kind of Hindi** she speaks, **what kind of Hindi** she speaks, **what kind [of Hindi]** I [speak].
> **Fig 87: Anita (F32), convo_B**

The dialogue occurred while one of the participants (Anita) was explaining the purpose of the research project to her landlady who had joined the group partway through the recording. In a culture where calling someone by name is not always considered polite a liberal use of pronouns is a good alternative; this is evidenced in the short passage in Fig 87 where third person singular pronouns are used to refer to the researcher (not present at the time) as well as other female research participants.[71]

Phrases formed with *type* and a demonstrative determiner are among the most common in the corpus, accounting for a third of all tokens (21/67). However, only a minority of these (three tokens) fulfil a categorizer function, the rest therefore showing grammaticalized usage. One example of categorizer usage occurs during a discussion about the teachers at a local school who (we learn) either move from class to class or work part-time and only come into school to teach their subjects. Anita describes the alternative, a full-time class teacher,

---

[71] The speaker used the proximal form of the pronoun *iski* to refer to me, although I was not present at the time.

and then (in Fig 88) asks her interlocutor to confirm that that kind of teacher does not teach the higher classes at the school.

hin	हाँ	इस	टाइप	के	नहीं	हैं	ना
tra	han	is	**type**	ke	nahi	hain	na
gls	yes	DEM	**type**	GEN.PL	NEG	AUX	TAG
eng	Yeah, they don't have that kind [of teacher], do they?						

Fig 88: Anita (F32), convo_b

The *type* phrase highlighted in Fig 88 contains a plural marking on the genitive, indicating an implicit plural category noun which has been dropped in a similar way to the example in Fig 86. The complete categorizer construction would contain the category noun *teacher*, as in *is type ke teacher*; however, as in the final token of Fig 87, the category noun *teacher* 'teachers' was explicit in the previous utterance and is therefore not needed.[72]

Because of the lack of category noun, this example can be considered a bridge to the similative demonstrative interpretation of the *type* phrase with the demonstrative; the motivation for this grammaticalization argument will be addressed in Chapter 9.

8.2.3   Manner descriptors in Hindi

In the previous chapter, I outlined the process by which type nouns cross-linguistically come to take on adverbial functions facilitated by both semantic shift and the application of adverbializers. This cross-linguistic tendency is borne out in Hindi, where adding the postposition *se* to adjectives and nouns produces an adverb of manner, as with Hindi nouns *aaraam* 'rest'→ *aaraam se* 'slowly, in a relaxed fashion' and *aasaani* 'ease' → *aasaani se* 'easily'.[73] This is just one of the many functions of highly polysemous grammatical morpheme *se*, also prolific as an instrumental marker. Type nouns *prakaar* and *tarah* have been recruited to form adverbial phrases, as evidenced in Figs 89 and 90.

> to **is tarah se** [sambandh] bana raha
> So **in this way** [the relationship] was formed and retained
> **Fig 89: Leela (F55), convo_a**

Fig 89 gives an example of a phrase with *tarah* in its manner sense, which with the proximal demonstrative and the adverbializer *se* which creates an adverbial phrase *is tarah se*

---

[72] The loanword *teacher* is unmarked in the plural, following the standard pattern for masculine nouns ending in a consonant.

[73] This is listed as sense 3 under से *se* in the OHED - 'in expressions of manner formed on nouns' (2012:1037).

meaning 'in this way'. The most common context in which *tarah* appears in the data as a manner noun is with the adjective *achha* 'good', inflected to agree with *tarah*'s grammatical gender (n.f.), as in *achhi tarah* 'in a good way'. Fig 90 shows a more metaphorical discourse structuring usage of the type noun *prakaar* to present an argument; *tarah* can also be found in this context.

> ***ek prakaar se*** *bhabhi hai teri vo*
> **In one way**, she's your sister-in-law!
> **Fig 90: Chandrashekar (M22), convo_e**

Finally, the similative demonstrative of manner, *aise*, also modifies verbs in a similar way to the adverbial phrase *is tarah se* and is by far the most frequent lexeme fulfilling this role. The similative demonstrative paradigm is a hugely important feature of the colloquial Hindi captured in the recordings from Pithoragarh and will be presented in detail in section 8.3. Here it will suffice to introduce the manner similative demonstrative *aise* 'like this, in this way', highlighted in Fig 91.

> *ham log* **aise** *baithhak me baithhe the*
> We (lit. us people) were sitting in the sitting room **like this**[74]
> **Fig 91: Mamta (F64), convo_b**

A further example (Fig 92) demonstrates the free word order which *aise* enjoys; although it traditionally sits before the verb, the speaker has chosen to emphasize the imperative here, displacing *aise.*

> *raat be raat bike me mat jaaya kar* **aise** *vahan*
> Don't go there by bike **like that** in the middle of the night!
> **Fig 92: Mohan (M32), convo_j**

Note that the adverbial phrase *is tarah se* could replace *aise* in the sentence in Fig 91, but it is unlikely to appear in the same postponed position as *aise* in Fig 92; it more commonly appears before the verb as we observed in Fig 89.

---

[74] Hindi lexeme *baithhak* has been translated using Br.Eng 'sitting room' to highlight its derivation from the verb *baithhna* 'sit'. The *baithhak* is a room in which guests are typically hosted and in a house with two living rooms will be the outermost of the two.

8.2.4 *Type* as a manner descriptor

In the Pithoragarh data, speakers also applied adverbializer *se* to loanword *type* and thus created an adverb of manner represented by the schema "X + TYPE + SE", 'in an X way', where X could be a determiner, a quantifier or an adjective. The meaning of this newly formed adverb requires an interpretation of *type* with the sense 'manner, way' and not 'kind, sort', a clear example of semantic shift. Fig 93 shows two examples of the adverbial use of the construction from convo_f.

hin	वो	उसी	टाइप	से	बोलेंगे	जिस	टाइप	से	उन्होंने	सीखा
tra	vo	usi	**type**	se	bolenge	jis	**type**	se	unhone	sikha
gls	she	that.EMPH	**type**	ADV	speak.FUT	REL	**type**	ADV	she	learnt
eng	She'll speak exactly the way she learnt (lit. in precisely that way, the way in which they learnt)[75]									

Fig 93: Monika (F29), convo_f

In this utterance, adverbial *type se* combines with distal demonstrative determiner *us* 'that' and corresponding relative adjective *jis* 'the one which' to form a relative-corelative construction, a common and productive construction in Hindi which facilitates the linking of two clauses through deixis. In Figs 93 and 94 the way a foreign woman is expected to speak Hindi is described as 'the way she learnt', using *type* as a manner descriptor. The use of the emphatic particle ही *hi* ('precisely, exactly') along with the demonstrative as *usi* 'exactly that' renders the meaning 'exactly that way'.

Monika	*kyonki vo **usi type se** bolenge **jis type se** unhone sikha.* *ham log to ek hi hindi ko **kitni tarah se** bolte hain na*
en	Because she'll speak **exactly the way** she learnt (lit. **in precisely that way**, **the way in which** she learnt). [But] we speak one Hindi **in so many ways**, don't we?

Fig 94: Monika (F29), convo_f

The *type* phrases *usi type se* and *jis type se* are used as the semantic equivalents of Hindi equivalent *usi tarah (se) and jis tarah (se)*. The adverbial usage was not prolific in the corpus (only three tokens from two speakers). Additionally, the speaker could have concluded her utterance with *type* in place of *tarah* in the phrase *kitni tarah se* 'in so many ways' but did not, perhaps suggesting that *type* does not yet collocate with quantifiers.

---

[75] Note that following North Indian culture, those of senior age or status are always referred to with plural morphology and pronouns.

There is a cross-linguistic precedent for type nouns to share both the categorizer sense of 'kind, sort' and the manner sense of 'manner, way'; however, it is not universal. In Russian, the polysemy is maintained by lexeme *rode* 'kind', but not *tip* 'type', whereas in Polish, type nouns do not hold a manner meaning (Kisiel and Kolyaseva, forthcoming; see also Kolyaseva, 2021a for Russian). Note that the OHED tends to list glosses in order of primacy not history and does not therefore indicate which of the meanings developed first in *prakaar* and *tarah*.

Anomalous cases

Despite a clear interpretation of *type* as 'manner' and clear uses of *type se* as an adverbial phrase, in three cases it seems that speakers do not distinguish between *type ka* and *type se*, but rather treat them as synonymous. In Fig 95, the phrase *puraane waale type se* 'older types of [flat]' appears in a discussion on the types of décor demanded of their rental accommodation by tenants in Pithoragarh. As a landlady, participant Mamta is considering redecorating one of her flats with tiles in order to make it more marketable.

> *ab **tiles** waale maangte hain jyaade. **puraane waale type se** nahi maangte hain*
> Now people ask for ones with **tiles** more. They don't ask for **kind of old [style] ones.**
> **Fig 95: Mamta (F64), convo_b**

Although preceding a verb, the adverbial reading *'old ones way of asking' does not make sense here, and the *type* phrase is clearly functioning as a modifier not of the verb but of the noun phrase *puraane waale*, 'old ones' (=old style flats), with the category noun 'flats' implicit from the discourse context. Here it seems the speaker is using *type se* as an alternative to *type ka* with an approximative meaning. (For a full discussion of hedging uses, see section 8.5 of this chapter.) This usage of the adverbializer *se* in place of the genitive *ka* is perhaps an indication that some aspects of the uses of *type* are not completely stable. Given its relatively short history in the region, this is not unexpected.

8.3 PART II Functions of *type*: phoric modifier and phoric similative

8.3.1 Phoric modifiers in Hindi

Chapter 7.2 described the functions and features of the phoric modifier, a binominal *type* construction common to many of the languages studied, as in *that type of cake* or *ce genre de voie* 'this kind of road' (example from Mihatsch 2016:151). Syntactically, this construction comprises a determiner phrase formed of a demonstrative determiner and the *type* cognate followed by a genitive which collectively precede (and modify) the category noun. Functionally, phoric modifiers facilitate the combination of instance and category reference. The presence of the demonstrative makes the modifier inherently phoric, although as we shall see, the vaguer the reference and the more unusual (or non-hierarchical) the category, the less truly anaphoric or cataphoric the phrase becomes.

Phoric modifiers in Hindi can be formed from the taxonomic noun *tarah*, although there was only one token attested in the data set (Fig 96).

> *das saalon me paintees officer mare hain, CRPF ke aur **is tarah ke forces** ke*
> in [the last] ten years, twenty-five officers had died, from CRPF[76] and from **those kinds of [armed] forces**
> **Fig 96: Suraj (M25), convo_m**

Similative demonstrative aisa

Much more frequently employed as a phoric modifier than the *tarah* construction is the complex lexeme *aisa*, whose lexical entry in the OHED (Fig 97) catalogues only part of its multifunctional nature.

ऐसा *aisa* **adj. & m.**

1. adj. of this sort
2. m. such a thing (as this)
3. ppn. like, similar to. मर्द ~, like a man
4. adv. in this way; as though. ऐसे ही, adv. just so, precisely; as if, as though; casually, thoughtlessly

---

Fig 97: Lexical entry for *aisa* from OHED

Classified as a similative demonstrative, a single lexeme which combines deictic reference with similative comparison (see Chapter 7.3), proximal *aisa* can fulfil the same semantic-

---

[76] Central Reserve Police Force, one of India's many armed forces separate to the Indian Army.

pragmatic functions as a phoric modifier phrase formed from *tarah* and a proximal determiner, as illustrated in Fig 98 where feminine-marked *aisi* could be substituted by the complex determiner *is tarah ki* lit. 'this kind of'.

> **aisi baat** *bhi nahi hai dimaag me na*
> **That kind of thing** isn't even in my mind, you know[77]
> **Fig 98: Bhuwan (M37), convo_a**

This distinction is similar to that of the dated English equivalent *such* and its more commonly employed synonym *like this/that*. Likewise, the distal similative demonstrative *vaisa* is synonymous with distal *us tarah ka* 'that kind of' in this context. Both *aisa* and *is tarah ka* function syntactically as modifiers of the following noun to which they both also show gender and number agreement. Additionally, both combine deictic properties with a dual reference to instance and type via ad hoc categorization and similarity, which are characteristic features of phoric modifiers.

The nature of similative demonstratives of quality is that they compare features or qualities of one object with those of another. This can involve reference to a similar type, although it is by no means the only reading. When functioning syntactically as a modifier, modifying an explicitly specified noun, it is therefore difficult (if not impossible) to separate categorizer readings from similarity readings; both are inherently and concurrently valid. In Fig 99 the categorizer reading is perhaps stronger, as evidenced by the preferred translation; however, a translation incorporating 'a problem like that', i.e. like the problem the interlocutor has just been describing, would also be accurate.[78]

> *achha, to ab to kuch nahi hoga, phir. abhi* **aisi problem** *nahi hui na*
> OK, so now nothing will happen, then. There isn't **that kind of problem** right now, is there?
> **Fig 99: Bhuwan (M37), convo_a**

When similative demonstratives appear in other syntactic contexts not immediately preceding a noun, particularly as a predicate in sentences with a copula (such as in Fig 100), the categorizer reading is loosened and the similarity reading foregrounded.

> *kamar bhi meri* **aisi** *ho gayi hai*
> My waist has also become **like this**
> **Fig 100: Bhuwan (M37), convo_a**

---

[77] Note that Hindi speakers use the proximal demonstrative to refer to previous discourse content, whereas in English speakers would tend to use the distal 'that', hence the apparent mismatch in the idiomatic translation. The question tag *na* is idiomatically translated as 'you know'.

[78] Note that 'problem' is an established loanword.

These non-modifier uses of the similative demonstrative, which will be classed as 'Phoric similative', are discussed in section 8.3.4.

Relative-corelative constructions

Elements of the quality content dimension function together in relative-corelative constructions such as that in Fig 101 where *aisa* partners with its relative counterpart *jaisa*. Note that both phoric modifiers combine a kind reading with a similarity reading as indicated by the translations, the literal translation favouring a kind reading and the idiomatic translation favouring a similarity reading. Both elements are inherently present.

hin	ऐसी	दुकान	खोल	लो	फ़िर	हमारे	जैसी
tra	**aisi**	dukaan	khol	lo	phir	hamaare	**jaisi**
gls	**SD.F**[79]	shop.F	open	INTF	then	PRO.1PL	**REL.SD.F**

eng   Open a shop **like** ours then! (lit. this kind of shop … the kind like ours)

Fig 101: Parvati (F57), convo_t

The example also demonstrates the high degree of flexibility in word order that is characteristic of colloquial spoken Hindi in general. The alternative construction would have been a similative usage of *jaisi* without *aisa*, rendering an utterance like the following: *hamaari jaisi dukaan khol lo* (lit. 'ours like shop open').

We have seen how similative demonstratives of quality fulfil a phoric modifier role modifying nouns. In the next section we will observe that the phoric modifier construction with *type* functions likewise.

8.3.2   *Type* as a phoric modifier

The structure of the phoric modifier with *type* is patterned on the demonstrative construction with *tarah* and has the same nuance of meaning, as Fig 102 shows. The context of this utterance is the participants of convo_g discussing conversations which one of them had overheard and all were displeased with. Urmila provides her commentary on the situation using hypothetical chastisement of the parties involved. She uses a phoric modifier with *type* which she repeats as if questioning the guilty parties. (Note the cooccurrence of

---

[79] Note that in the glossing, SD with no additional marking represents the proximal. The prefix DIST will be applied to the distal form and REL for the relative adjective/pronoun form.

the similative demonstrative of manner *aise* 'like this', which was introduced in the previous section*.)*

> *vaise kaho to kitni galat baat hui. tum aise **is type ki baate** karte ho. kyon karte ho **is type ki baate**?*
> If you put it like that then it's really a bad thing! "You're talking about **these kind of things**, like this. Why are you talking about **these kind of things**?"
> **Fig 102: Urmila (F47), convo_g**

The bold text indicates the *type* phrase with the category noun underlined: the structure of the phrase appears equivalent to the categorizer usage. Fig 103 presents the final sentence fully glossed to facilitate more detailed analysis.

hin	क्यों	करते	हो	इस	**टाइप**	की	बातें
tra	kyon	karte	ho	is	**type**	ki	baate
gls	why	do	AUX	DEM	**type**	GEN.F	things said.FPL
eng	Why are you talking about **these kind of things**?						

Fig 103: Urmila (F47), convo_g

The schema can be represented as DEM + TYPE + GEN + N. The demonstrative complex with *type* fulfils the same function as *aisa*: it defines one subtype (or subcategory) of the specified category on the basis of properties of the (deictically referenced) instance. Unpacking this statement with reference to the example in Fig 103 (*is type ki baate*) yields the following:

1. The specified category is *baate* 'things said'[80]

2. The deictically referenced instance (*is* 'this') refers to the 'things said' which are available in the "discourse memory" (recent discourse context)

3. The *type* (subcategory) is therefore defined as an ad hoc category or set of 'things said' which includes those things mentioned in prior discourse but can also include other similar items.

This definition of the phoric modifier differs slightly from that of Mihatsch (2016:151-2), who describes the demonstrative complex as based on a relation of identity (with the instance) and not similarity. The position adopted here is that both relationships are supported by the construction: identity with the named (deictically referenced) instance and similarity with other possible members of an ad hoc group conjured up by the mention of a 'type'.

---

[80] बात *baat* is an incredibly versatile word whose first sense alone is glossed in the OHED as 'something said, a word, remark; speech, talk, words; conversation; discussion'. In the sentence in Fig 107 *baat* forms part of a light verb *baat karna* 'talk' (lit. do words), which is split for expressive purposes with the verbal 'do' foregrounded. A literal translation of the sentence would read, "Why are you doing these kind of words/talks?"

Why would a speaker wish to use a postdeterminer complex (or a similative demonstrative modifying a noun) rather than a simple demonstrative, as in 'these things'? Leaving aside sociolinguistic issues related to the speech style of a community and the way things are usually said, there are other motivations. One such motivation is the appeal to a larger set; the dual reference, the juxtaposition of the immediate ('this') and the generic (*type*), extends the reach of the discussion beyond the immediate and envelops other possible things that have been said or might be said. At the same time, the ambiguity of the dual reference (being neither completely one nor the other) lends itself to vagueness, which has numerous discourse-related benefits, including mitigation for politeness purposes.

The example below (Fig 104) is pragmatically even vaguer than Fig 103 by virtue of a very loose deictic referent for the 'sort of man' in recent discourse context and syntactically the separation between the demonstrative modifier and the category noun *aadmi* 'man' brought about by the insertion of the contrastive particle तो *to*.[81]

hin	इस	**टाइप**	का		तो	आदमी	वह
tra	is	**type**	ka		to	aadmi	vo
gls	DEM	**type**	GEN.MSG		EMPH	man	PRO.3SG

eng   He's just that sort of man

Fig 104: Mohan (M32), convo_j-2

All three examples of the phoric modifier with *type* from the Pithoragarh data include negative connotations, a nuance shared by *tarah* and *aisa*. The vagueness of the deictic reference allows a speaker to be negative without being completely specific, thereby saving face. This usage foreshadows the phoric similative function where no category is specified (although a general category like 'person' could perhaps be imagined), rendering a sentence such as 'S/he's just like that'. The phoric modifier construction is therefore one possible bridging site from nominal to non-nominal (phoric similative) uses of *type*.

---

[81] For a thorough overview of the many uses of the contrastive particle *to* in Hindi, see Montaut (2015).

8.3.3 Introduction to similative demonstratives in Hindi - the quality/manner distinction

Very little has been written about similative demonstratives in Hindi. Given what a vital role they play in discourse, this is perhaps surprising. What follows will be a brief introduction to the form and function of these highly flexible lexemes.

The table in Fig 105 shows the proximal and distal forms of all Hindi similative demonstratives as well as their corresponding relative pronouns and interrogatives. The table presents the base form of the word which in the case of quality and degree demonstratives is inflected for gender and number.

Proximity Dimension / Content Dimension	Proximal	Distal	Relative	Interrogative
Quality	*aisa*	*vaisa*	*jaisa*	*kaisa*
Manner	*aisa/aise*	*vaisa/vaise*	*jaise*	*kaise*
Degree	*itna*	*utna*	*jitna*	*kitna*

Fig 105: Similative demonstratives in Hindi based on categories in König (2017)

On the proximity dimension, Hindi recognizes proximal and distal determiners including the proximal similative demonstrative of quality *aisa* and its distal counterpart *vaisa*. Additionally, it has a derived relative adjective and pronoun (*jaisa*) and an interrogative (*kaisa*), as outlined in Sahoo and van der Auwera (2019:29, Table 4). Hindi prefixes are common across all similative and demonstrative forms (including simple demonstrative determiners) and involve the prefixation of व- *v-* for the distal form, ज- *j-* for the relative adjective/pronoun form and क- *k-* for the interrogative form. These prefixes are also shared by Bengali and Odia (Sahoo and van der Auwera, 2019).[82] As with many nouns and adjectives in Hindi, the masculine singular form ends in आ /aː/ *-a*, and is inflected with ई /iː/ *-i* for feminine singular and ए /eː/ *-e* for plural.[83]

On the content dimension, spoken Hindi differentiates between quality, manner and degree demonstratives, although the dictionary (OHED) seems to be lagging behind. Only *aisa* and not *aise* has its own lexical entry in the OHED and manner adverbial functions of *aise* are

---

[82] The distal degree demonstrative is irregular (*utna* not **vitna*) but retains the bilabial phonetics shared by consonant व- *v-* and vowel उ- *u-*.

[83] Although standard Hindi uses an *-i* ending for feminine plural (the same as feminine singular), in the Pithoragarh variety of Hindi, the masculine plural ending *-e* is used for both grammatical genders.

included within the entry for *aisa*.[84] However, the data from the corpus validates their separation: only 43/227 tokens of phoric similative and similative manner usage were not fulfilled by lexemes inflected with the adverbial morpheme *-e (*namely *aise, jaise* or *vaise*).

As discussed in the previous chapter, South Asian languages do not typically distinguish between quality and manner (see Sahoo and van der Auwera, 2019), but a distinction between the two categories in Hindi was (perhaps unintentionally) suggested by Montaut (1995), who presents what is described in this thesis as the manner category as simply an adverbial form of the quality lexeme (see the example in Fig 106), used whenever the comparator is non-nominal.

> *uske pas **jaisa khilauna** hai, **vaisa hi khilauna** tumhe milega*
> You will get a toy exactly like his/hers
> (lit. the sort of toy s/he has, that (only) sort of toy you will get)
> **Fig 106: Copied from Montaut (1995:158, footnote 3)**

Although data from Japanese and other languages (König, 2017) suggests that manner demonstratives are the most basic or primary of the three, from which the others are derived, in Hindi there appears at face value to be an argument for considering *aisa*, the proximal similative demonstrative of quality, as the basic one. Indeed, it is the most frequently occurring of all the similative demonstrative lexemes in the corpus (N=801). The hypothesis proposed here is that similative demonstrative *aisa* originally covered both manner and quality contexts, but over time and due to the influence of a productive process of adverbialization in Hindi using (the adverbializer) *se*, the plural form *aise* developed as a new and distinctive similative demonstrative of manner, along with its corresponding distal form *vaise* and relative adjective *jaise*.

Fig 107 is a relative-corelative construction where both elements are pronominal similative demonstratives.

---

[84] Another plausible influence on the development of a new adverbial form is via the plural form of *aisa* (*aise*), now homophonous with the manner similative demonstrative, or more specifically via its interrogative form *kaise*. Additionally, the interrogative adjective and adverb meaning 'how? why?' is listed in the dictionary as *kaisa* although the manner form prominent in colloquial speech is incontrovertibly *kaise*, as illustrated in the following examples: plural form *uncle ji **kaise** hain?* '**How** is respected uncle?' and manner interrogative *us ne **kaise** kiya?* '**How** did s/he do it?'

hin	जिसने	**जैसा**	कहा	**वैसा**	कर	दिया	मैंने
tra	jisne	**jaisa**	kaha	**vaisa**	kar	diya	meine
gls	who.ERG	**REL.SD.PRO**	say.PST	**DIST.SD.PRO**	do	INTF.PST	1SG.ERG
eng	I did what everybody told me to do. (lit. who that-which said like-that did I)						

Fig 107: Mamta (F64), convo_b

All members of the similative demonstrative paradigm displayed in the table in Fig 105 are highly polysemous and show evidence of having taken on new grammatical (and semantic-pragmatic) meanings.

8.3.4   Phoric similatives in Hindi

In this section, similative comparison within Hindi is introduced, providing a framework for a discussion of how *type* constructions have also come to function as similatives, first investigating deictic comparison (comparing an object or person who is 'present' either physically or in discourse with one who is merely described) and then non-deictic (comparing two objects or people with each other). The two kinds of similatives are not often discussed or analysed together in the literature. Yet the present context, not only of similatives in Hindi but also the innovative extended uses of *type* in Hindi, demands such treatment.

Phoric similatives can be differentiated from phoric modifiers in terms of the greater vagueness of their deictic reference and the preference for a similarity reading over a categorization reading (although vague categorization is still present). In other words, a phoric similative may claim that someone is 'like that' while not describing exactly what they are like, leaving the hearer to infer characteristics or features of the person on the basis of (perhaps) a single anecdote. An example of this is in Fig 108 where Ankita is describing a colleague's actions and extrapolates these to a claim about the colleague's character using the distal similative demonstrative of quality *vaisa*, which is inherently vague in this (syntactic) context: 'She <u>is</u> like that'.

> *Preeti kaise gussa ho rakhi hai mujhse! gussa kya pata?! hai **vaisi** vo, **us type ki***
> How angry has Preeti got with me! Who knows [if she's] angry?! She <u>is</u> **like that**, **of that sort / like that**
> **Fig 108: Ankita (F28), convo_d**

The *type* phrase that follows (*us type ki*) is synonymous with *vaisi*, the translation highlighting the dual reference of (vague) categorization and similarity which is a feature of similative demonstratives of quality. Phoric similatives commonly occur as adjectival predicates with the copula in a structure like 'SUBJ + (V)AISA + COP', although pronominal uses are also attested (see below).

While the previous example was anaphoric, the example in Fig 109 is exophoric – the researcher was showing the participants photos of motorbikes on her phone and Viren was incredulous about the way one of them looked.

> *ye bike bhi hakikat me hogi ye - **aisi**? **aisi** to kya hogi vo!*
> Is this bike also for real - **like this**? It can't be **like this**!
> **Fig 109: Viren (M20), convo_h**

The word order is expressively arranged so that similative demonstrative *aisa* (with feminine inflection for *bike*) appears utterance-finally in the first sentence, and in focal position in the second sentence; in both cases, the similative demonstrative is an adjective and (along with the copula) a predicate of the subject *bike*. The lack of category noun ensures that the similarity reading is paramount, although with adjectival similative demonstratives, the vague sense 'of this sort' is always retrievable.

The relative similative demonstrative *jaisa* partners with either *aisa* or *vaisa* to form a relative-corelative sentence with one of each of the pair in each clause. The following example shows how not only proximal but also distal and relative similative demonstratives can be used together in relative-correlative constructions functioning as phoric similatives.

hin	जैसी	[दिल्ली][85]	की	लड़कियाँ	होती	हैं	वैसी	नहीं	है
tra	**jaisi**	[Delhi]	ki	ladkiyan	hoti	hain	**vaisi**	nahi	hai
gls	**REL.SD.F**	[Delhi]	GEN	girls	be.FPL	AUX.PL	**DIST.SD.F**	NEG	AUX
eng	She isn't **like** [Delhi] girls (lit. **of such a sort as** girls of Delhi are, **of that sort** not is)								

Fig 110: Ankita (F28), convo_d

There are two possible translations of Fig 110, one idiomatic and one literal. The literal translation of Fig 110 is based on the dictionary definitions of *jaisa* and *vaisa* and emphasizes the semantic sense of categorization. However, there is a key difference: the categorization is vague. In the absence of a category noun, the vague type reading 'of that sort' is reinterpreted as a similative reading 'like that', which takes precedence.

Pronominal uses 'something like that; that kind of thing'

It seems that as *aise* became differentiated as an adverbial manner similative demonstrative, *aisa* was reanalysed as a pronoun when paired with a verb, thus reinforcing its status as a similative demonstrative of quality. In this context (alongside a transitive verb), the similative demonstrative has acquired pronominal status. The context of Fig 111 is three teenage girls

---

[85] Location changed because of sensitivity.

(16+) discussing their evening out the day before and congratulating themselves that they would not have to regret having fun. The sentence in Fig 111 is hypothetical speech expressing the regret the girls would not have to feel. Similative demonstrative *aisa* functions as a pronominal placeholder for the fun they had had.

> *"kaash hamne **aisa** kiya hota"*
> "If only we had done **something like that**"
> **Fig 111: Neelima (F16), convo_n**

Although *aisa* and its related lexemes are by far the most prolific of all phoric similatives, mention must also be made of extended uses of categorizer and manner noun *tarah* which appears in phrases with the demonstrative in contexts where *aisa* could also appear, and with a similar meaning. One such example is in Fig 112 where Suraj is describing a senior member of the Indian civil service who ruffled people's feathers standing up to corruption and bad practice.

> *to vahan bhi usne bhanda fod diya, matlab hai hi banda **is tarah ka***
> So there too he busted a racket. I mean, the guy is just **like that** / just **that sort**
> **Fig 112: Suraj (M25), convo_m**

The phoric demonstrative phrase *is tarah ka* is employed (in the absence of a category noun) as part of the predicate describing the *banda* 'guy' as similar to the sort of person who would perform actions such as those mentioned previously *usne bhanda fod diya* 'he busted a racket'. In the absence of a category noun, the categorization is inherently vague, and the standard must be understood as the same as the comparee, the sentence subject *banda* 'guy'.

Extended uses of the phoric similative

As a phoric similative, *aisa* shows evidence of subjectification, one of the features of a grammaticalization process. As *aisa* grammaticalizes it is prone to increasing vagueness which can be used pragmatically for euphemism, to convey negative connotations, or to avoid answering a question directly, particularly in combination with the emphatic particle *hi* 'only' whereupon it is rendered *aisi*. The adverbial usage meaning 'carelessly; thoughtlessly' is dictionary attested (see Fig 97), but the pronominal usage is not, which suggests that it is a more recent development.

The more extreme end of the vagueness continuum of *aisi* usage is epitomized in a conversation between two young men Viren and Lokmanyu (Fig 113). Viren, being well-known to the researcher and wanting to help out, was trying to get his friend to talk for the

recording, but his friend was apparently not in a chatty mood, repeatedly returning his invocations with the ultra-vague response *aisi hai* 'the same as always, nothing to speak of'.

#	Speaker	Utterance
1	Viren	*aur, aur kuch suna yaar*
	en	And? Tell me something more, *yaar*[86]
2	Lokmanyu	*aisi hai abe*
	en	It's **just like that (=nothing to speak of)**, *abe*![87]
3	Viren	*"aisi hai" bolega na*
	en	You will say, "It's **just like that**", won't you?

Fig 113: Viren (M20) and Lokmanyu (M19), convo_h-2

Viren's frustrated final remark highlights the fact that the use of *aisi* communicated nothing of any content or consequence.

8.3.5 *Type* as a phoric similative

Type phrases with the demonstrative determiner (both proximal and distal) account for nearly a third of all *type* tokens. This reflects both the dynamic nature of colloquial discourse and, as shall be seen in this section, the innovative uses to which demonstrate phrases with *type* have been put in Pithoragarh Hindi.

Phoric similative constructions constitute 8 of the 67 tokens of *type* in the Pithoragarh data. They appear in a variety of syntactic contexts, as predicates in existential sentences, as adjectival modifiers of both sentence subjects and objects and also as pronominal sentence subjects. (Note that these are exactly the syntactic contexts in which *aisa* and its related lexemes also appear). Fig 114 is a glossed version of the final sentence of Fig 108 and illustrates how the similative demonstrative lexeme *vaisi* (in the red box) has an equivalent meaning to the complex *type* phrase *us type ki* (in the blue box). Both are examples of adjectival (predicative) usage.

hin	है	वैसी	वो	उस	**टाइप**	की
tra	hai	vaisi	vo	us	**type**	ki
gls	COP.3SG	DIST.SD.F	3SG	that	**type**	GEN.F

eng She is **like that**, **like that** (lit. that sort of)

Fig 114: Ankita (F28), convo_d

---

[86] *Yaar* is a Hindi word literally meaning 'friend', but is used in conversation similarly to *man* or *dude* in Br/Am. Eng.

[87] *Abe/Abey* is a Hindi swear word.

The two similative demonstratives, one simple (*vaisi*) and one complex (*us type ki*), are semantically equivalent, combining a type reading and a similarity reading, as highlighted earlier. It is intriguing therefore that both were used together in repetition. The sentence word order has been innovatively manipulated to emphasize the copula *hai* 'is', and using *us type ki* as an utterance-final tag provides emphasis. Additionally, the taxonomic meaning would be spotlighted by the explicit mention of *type* (an example of semantic persistence), adding a categorizing nuance to the phrase and prompting its use for expressive purposes.

The feminine marking on the genitive in Fig 114 matches the inflection on the distal similative demonstrative *vaisi*, both indicating that the subject of the sentence is a female. In this example, *type* does not exhibit nominal properties and the genitive, while still present, cannot agree with the noun which follows it (because there is none), but rather agrees with the gender and number of the subject. The *type* phrase *us type ki* thus appears to have become an indivisible unit that can be analysed as a similative demonstrative (of quality). Here we must conclude that *type* has incontrovertibly lost its nominal status. The phoric similative therefore constitutes the first non-nominal usage of *type*.

There are also cases where an adjectival phrase modifies not the subject of the sentence but the object, such as in Fig 115 where two neighbours are looking at a Facebook post in which someone claims to have seen (unidentified) red and green lights in the sky. As they are discussing the comments that others have made on the Facebook post, one neighbour tells another that something like this has also been seen in their town. The demonstrative *type* construction *is type ka* 'like this / of this sort' thus refers deictically to a specific discourse proposition – red and green lights.

Anita	*jo abhi upar aayi thi to logon ne kya likh rakha hai uske baare me, han?*
en	The thing which came above just now, what have people written about it, yeah?
Anita	*pithoragarh ke ek ne kuch dekha shaayd -* **is type ka**, *aakaash me.*
en	Someone from Pithoragarh saw something, I think - **like this**, in the sky.

Fig 115: Anita (F32), convo_b

A pronominal usage of the phoric similative is exemplified in Fig 116, where the higher prosodic pitch and emphasis on the word *ye!* 'This!' immediately preceding the *type* phrase helps us imagine Preity showing her spots to her neighbours:

ye! **is type ka** *ye. khujli se, na,* **is type ka** *ho raha tha*
This! **Like this** here! Through itching, you know, **such a thing (as this)** was happening

Fig 116: Preity (F42), convo_b

The second sentence is repeated with glossing in Fig 117. The *type* phrase, a unit equivalent to synonym *aisa*, is the only possible sentence subject.

hin	कुजली	से	न	इस	**टाइप**	का	हो	रहा	था
tra	khujli	se	na	is	**type**	ka	ho	raha	tha
gls	itching	INS	Q TAG	DEM	**type**	GEN	happen	PROG	AUX.PST

eng   From itching, you know, **such a thing (as this)** was happening.

Fig 117: Preity (F42), convo_b

It seems likely that the development of the demonstrative *type* phrase as a pronoun as well as an adjective is modelled on the grammaticalization *aisa* and occurred in reverse order (adjective first). This cannot be confirmed without diachronic data, but tentative arguments supporting this theory will be presented in Chapter 9.

8.4   PART III Functions of *type*: similative

8.4.1   Similatives (of quality) in Hindi

In the previous part of the chapter the lexico-grammatical category of the similative demonstrative in Hindi was introduced and its functions as a phoric modifier and phoric similative were illustrated. The example in Fig 110 showed how corelative *jaisa* can function together with *(v)aisa* to explicitly compare two objects, people or ideas, one visible (exophoric) or known from recent discourse (endophoric) and one to which it is being compared. *Jaisa* also functions independently from *aisa* as a complex postposition linking two objects and facilitating the comparison of one with another. It is likely that this usage developed from the corelative usage, but a diachronic study is needed to confirm this. In this section, the function of similatives in Hindi and similative comparison is presented and similatives *(ke) jaisa, ki tarah* and *ki prakaar* are illustrated before the use of *type* as a similative is investigated. Focus will be given to similatives of quality not manner, since loanword *type* does not function as a similative of manner.

Jaisa as a similative (quality)

The most prolific similative in Hindi is *jaisa*, introduced in the previous section as the relative adjective and pronoun of similative demonstrative *aisa* (see Fig 105 for the full paradigm). The lexical entry for *jaisa* in the OHED (Fig 118) recognizes this primary sense alongside two others, including the postpositional usage with a similative meaning as secondary. Dictionary recognition of the fact that *jaisa* has grammaticalized and taken on a new function and a

corresponding new form suggests that the similative use of the lexeme has been established for some time.[88] Note that the lexical entry highlights *jaisa*'s role as a phoric modifier (#1) and similative (#2) but omits uses of *jaisa* as a hedge and an incipient quotative; these 'newer' functions will be detailed in future sections.

जैसा *jaisa* **rel adj. & pron.**

1. of such a sort as
2. (as ppn.) like, similar to. उन जैसी औरतें, *un jaisi aurate*, women like them
3. that which; the kind of… which. - ~ कि, **conj.** just as; as though

---

Fig 118: Extracts from OHED definition of जैसा *jaisa*

The example in the lexical entry for sense #2 of *jaisa* in Fig 118 (*un jaisi aurate* 'women like them') uses the schema "N + JAISA + N", where the first nominal slot was actually filled by a pronoun (the third person plural with oblique marking, *un*). The phrase in Fig 119 from the Pithoragarh data follows the same pattern. To use terminology adopted in the volume on similatives by Treis and Vanhove (2017), *jaisa* is the <u>standard marker</u> and as such always directly follows the <u>standard</u>, here the town Pithoragarh. In this structure the <u>comparee</u> follows *jaisa* creating a complex noun phrase which can function as a sentence subject or, as in Fig 119, an object. Note that *jaisa* agrees with the comparee in gender, here feminine *jagah* 'place'.

> *aur **pithoragarh jaisi jagah** to har kisi ko pasand aati hai*
> And everybody likes **a place like Pithoragarh** (lit. Pithoragarh like place) [89]
>
> **Fig 119: Mohan (M32), convo_j-1**

Hindi postpositions are commonly formed with the genitive marker, such as *ke liye* 'for', *ke baavjud* 'in spite of', *ki taraf* 'in the direction of', and the similative of quality and manner *ki tarah* 'like'. The example from the dictionary entry in Fig 118 (*un **jaisi** aurate* 'women **like** them') illustrates the way in which postposition *jaisa* despite lacking the genitive marker nevertheless causes the oblique inflection of the standard, in this case the third person plural pronoun उन *un* 'them' (which would otherwise be uninflected वे *ve*). There are five similar

---

[88] This observation will become important as we progress through the chapter and provide evidence of further grammaticalized uses, particularly as a hedge. Dictionary updates are necessarily retrospective and therefore appear later than in colloquial speech.

[89] The meaning of reflexive verb *(kisi ko) pasand aana* in Hindi is of the feeling of 'liking' or appreciation coming to someone upon their first experience of something. Here Mohan cannot believe that the new wife will not immediately like Pithoragarh because <u>everyone's</u> first impression of the town is positive.

examples in the data, in which the pronoun is inflected in the oblique as a result of the presence of *jaisa*.

However, the lexical entry does not list a form of the postposition with the genitive present (contrary to the lexical entry for *tarah*, where the similative sense is seen to require a genitive marker; see Fig 8, sense 5). It is curious, therefore, that the data from Pithoragarh exhibits 16 (out of 81) tokens of similative *jaisa* with an explicit genitive, such as Fig 120, which appears in a discussion centred around business opportunities and the popularity of homemade food. Jatin comments on how informal roadside restaurants which serve the local potato dish *gutka* are attracting customers because the quality (taste) of the dish is similar to the homemade version. Here we see the postposition *jaisa* appearing with its genitive. The construction identified can be summarized with the schema "N + GEN + JAISA + N".

> *kuch log bana rahe hain **ghar ka jaisa gutka***
> some people are making **gutka like home[made]** (lit. home GEN like *gutka*)
> **Fig 120: Jatin (M47), convo_t**

One explanation for the presence of the genitive is overgeneralisation by speakers familiar with other complex postpositions, although the agreement of both number and gender evident from the examples in the corpus show a distinct difference from other complex postpositions which employ invariable forms of the genitive (either the masculine plural *ke* or the feminine *ki*).

In addition to the constructions with *jaisa* flanked by two nouns illustrated above, *jaisa* also appears with only the comparison standard and a copula. Constructions with the copula constitute syntactic contexts known for ambiguity cross-linguistically; they are highly flexible, and in Hindi as well as many other languages accept both adjectives and nouns as predicates. In Fig 121, Tarun and Rashmi are discussing their granddaughters and exclaim that one of them looks like her dad, their son Pankaj. The similative comparison is achieved using the 'standard' *Pankaj* and the 'standard marker' *jaisi*, which agrees with the sentence subject ('comparee'). This construction can be summarised as "SUBJ + N + JAISA + COP".

> *ye hain hamaari naatniyan. kaisi lag rahi. ye alag hi alag hai. ye **pankaj jaisi** hai*
> Here are our granddaughters. How do they look? She is totally different. She's **like Pankaj**.[90]
> **Fig 121: Rashmi (F53), convo_k**

---

[90] Names in all examples have been changed to preserve anonymity.

Fig 122 is an example of the same similative construction with *jaisa* but a different word order. We can see that the standard marker always follows immediately after the standard, regardless of the position of the comparee, which in this case is demoted to an utterance-final position.

hin	फ्रिज	का	पानी	**जैसा**	लगता	है	नौले	का	पानी	ना
tra	fridge	ka	paani	**jaisa**	lagta	hai	naule	ka	paani	na
gls	fridge	GEN	water	**like**	taste.PRS	AUX	*naula*	GEN	water	Q TAG

eng: The water from the *naula* tastes **like water from the fridge, doesn't it?**[91]
(lit. *naula* water … fridge water)

Fig 122: Tarun (M59), convo_k

The orange arrow indicates where the sentence subject would be placed had 'water from the fridge' not been topicalized and moved to the beginning of the utterance. The sentence structure can therefore be analysed as in Fig 123.

naule ka paani	***fridge ka paani***	***jaisa***	*lagta hai*	*na*
COMPAREE	STANDARD	STANDARD MARKER	VERB	Q TAG

Fig 123

We observe that in contrast to the structure in Figs 119 and 120 where *jaisa* is flanked by the standard and the comparee, the comparee in Fig 123 is relatively independent of the standard and its marker. The reason for this is the difference in relationship between the standard and the comparee. In the "N + JAISA + N" structure, the relationship between the two nouns is often one of subcategorization: the standard is a hyponym of the comparee (the hypernym), as in *Pithoragarh jaisi jagah* 'a place like Pithoragarh' where the standard *Pithoragarh* is a hyponym of the broader noun *jagah* 'place'. The comparison could also be interpreted as ad hoc categorization in which Pithoragarh is named as an example of the category leading to a sense which could be paraphrased 'a place **of a similar kind to Pithoragarh'**. Conversely, in Fig 123, the standard and the comparee are both equivalent, both hyponyms of the same hypernym, water. The nature of the similative comparison undertaken by this construction is therefore also different and the lack of categorizing undertones suggests it is the later developed structure. The loosening of the relationship

---

[91] *Naula* is the Kumaoni word for बावली *baawali* a 'small, deep tank with steps' (OHED) constructed by villagers to store slowly running water from a natural source.

between the standard and the comparee is a likely bridging context to the hedging function in which the comparee disappears, leaving *jaisa* to mark the 'standard' with a new emerging meaning as an approximative (see section 8.5).

Tarah as a similative

In addition to *jaisa*, another Hindi similative which was introduced earlier in the chapter is the categorizer noun *tarah* (see section 8.2.1). Forming a postposition with the genitive (*ki tarah*), it appears in a similar structure to *jaisa*, as illustrated in Fig 124, where *uski personality* 'his personality' is compared with the iconic motorbike Bullet. The speaker does not justify his claim of similarity between his friend and the bike or specify which features are similar; the hearer is left to reach these conclusions.

> *jo uski personality hai vo **Bullet ki tarah** hai*
> His personality is **like a Bullet**[92]
> **Fig 124: Chandrashekar (M22), convo_e**

This is the only token in which *tarah* functions in this way as it seems to be specializing as a similative of manner, a function not fulfilled by *type* at all. The distribution of functions between the different lexemes is an important subject which will be examined in the following chapter.

It should be mentioned that *jaisa* also functions as an equative in certain contexts: always when collocated with cardinal numeral *ek* 'one' as in *ek jaisa ghar* 'the same [kind of] house' and also whenever equity can be conceived of by implication, such as when Usha is describing the quality of her traditional Punjabi aubergine dish *bharta* in Fig 125.

> *tasty banti* [sic] *hai vo to bahut! **mere jaisa bharta** kaun banaayega*
> It comes out really tasty, that! Who can make **bharta like mine** (=as good as mine)?
> **Fig 125: Usha (F59), convo_q**

It seems that collocation with a personal pronoun can often lead to an equative reading. In another recording (convo_a), Anita discovers that her uncle suffers some of the same health complaints as she does and concludes: ***mera jaisa** hai aapko!* 'You've got **the same as me**!'

8.4.2   *Type* as a similative (of quality)

In addition to functioning as phoric modifiers and phoric similatives, *type* constructions are also used as (non-phoric) similatives, rendering a meaning of 'like, similar to'. There are two

---

[92] The Bullet is an iconic model of Royal Enfield motorcycle manufactured in India.

main syntactic constructions in which similative *type* appears, both mirrored by its Hindi equivalent counterpart *jaisa*.

The less common construction is one which explicitly references the second noun (Y). There is one such example in convo_p-2, where two women (Riya and Babita) are discussing the hair of another neighbour/relative. The complete extract is presented in its glossed form in Fig 126. The utterance contains a *type* construction akin to "N + TYPE GEN + N". The first noun is a Kumaoni loanword *dokka* which denotes a particular style of woven basket used for transporting dung, grass and other farm-related material from place to place in the village.

hin	डोपका	**टाइप**	के	बाल	हुए	उसके	ऐसे-ऐसे	घुँघरालू	करके
tra	dokka	**type**	ke	baal	hue	uske	**aise-aise like this x2**	ghunghraalu	karke
gls	basket	**type**	GEN.PL	hair.PL	became	her	**like this x2**	curly	do. ADV

eng   Her hair became **like** a *dokka* basket - **like this and this** - made curly

Fig 126: Riya (F39), convo_p-2

In describing her friend's hair after her perm as *dokka type ke* 'like a *dokka* basket', Riya is using *type* as a similative and through metaphorical extension presenting the shape of the friend's hair as similar to the shape of the *dokka* basket. Since the recordings were audio only and not video, it is not possible to ascertain whether the speaker was gesticulating; however, this is plausible and would be expected from the context. The image in Fig 127 shows a Kumaoni woman with a *dokka* basket – it is bigger at the top and then tapered towards the bottom.

Fig 127: Photo of a Kumaoni woman carrying a *dokka* basket

The *type* construction in Fig 126 contains the genitive marker which agrees with the following noun 'hair', a countable noun in Hindi.

The syntactic pattern "N + TYPE + GEN + N" illustrated above is less frequent among similative uses than the reduced construction with only the standard and the standard marker (*type ka*) which can be expressed as "N + TYPE + GEN". In the reduced construction the genitive *ka* remains, although it is redundant as a genitive (because there is no noun following it); with its new clitic-like status, its one role is to show inflection, similarly to *jaisa*.

Fig 128 is an excerpt from the first two minutes of recording h-1. Two friends are sitting together and Lokmanyu is keen to do something more interesting than 'just' record a conversation. However, he was evidently still aware of the recording device and the purpose of their getting together, and in between trying to get access to his friend's mobile to watch a film and chatting about a mutual friend, Lokmanyu suddenly makes this self-disclosure about his Hindi dialect.

#	Speaker	Utterance
1	Lokmanyu	*hamaari bhaasha kaisi hai pata hai?*
	en	Do you know what our Hindi is like?
2	Lokmanyu	*nepaali **type** ki hai*
	en	It's **like** Nepali
3	Lokmanyu	*hindi bhaasha ham **nepaali jaisa** bolte hain*
	en	We speak the Hindi language **like Nepali**

Fig 128: Lokmanyu (M19), convo_h1

Lokmanyu completes his revelatory disclosure by repeated his assertion in a different way, using *nepali jaisa* 'like Nepali' as a synonym for *nepali type ki* 'like Nepali'. Repetition of a phrase uttered in one language in another language for the purpose of clarification or emphasis is a common phenomenon in bilingual discourse (Gardner-Chloros, 2009:75).

We have observed that *type* is (almost) always followed by the genitive marker. In the previous section, I highlighted anomalous cases where the genitive marker appears before *jaisa*. There is also a single *type* token in the corpus, shown in Fig 129, which has a genitive both before and after.

hin	दुश्मन	के	**टाइप**	का	चेहरा	था	वो
tra	dushman	ke	**type**	ka	chehara	tha	vo
gls	enemy	GEN	**type**	GEN	face	AUX.PST	3SG

eng   It was a face **like** that of an enemy.

Fig 129: Urmila (F47), convo_g

The topic of conversation in Fig 129 is an old woman who was on her deathbed. The speaker claims to have known that she was about to breathe her last, and she describes the woman's

face in a number of ways, including this description of her face as 'black… **like** that of an enemy'. The use of a genitive *ke* before *type* as well as the genitive (*ka*) after seems to imitate the overuse of the genitive before Hindi postposition *jaisa* which we have already established is not required, but most likely an overgeneralisation of the normal pattern for (complex) postpositions.

8.4.3   Similatives (of manner) in Hindi

Both *jaisa* (in its adverbial form *jaise), tarah* and *prakaar* function as adverbial similatives of manner, in order of frequency. Montaut (1995) claimed that adverbial similative comparison was more frequent than nominal comparison. Yet in the Pithoragarh data set the reverse is true, with tokens of similative of quality from *aisa* and related lexemes (N=85) nearly double those of similative of manner (N=44). Similatives of manner incorporate both relative-correlative constructions and constructions using solely relative adverb *jaise*.

> *to **lawyer ki tarah sochna** padega aapko yahan par thoda sa*
> So you will have to **think like a lawyer** here, a little bit
> **Fig 130: Suraj (M25), convo_m**

In Fig 130 the meaning of *like* in the phrase 'think like a lawyer' means 'think **the way** a lawyer thinks' or '**in a similar manner** to a lawyer'. The same meaning could have been achieved using *ke jaise* or, more likely, a relative-correlative sentence using *jaise… vaise* 'in the manner in which… in that same manner'. An example of a construction with *jaise* fulfilling a similative of manner function from the spoken data is in Fig 131. Although the correlative adverb *vaise* is not spoken, it is implicit in the structure and meaning of the utterance.

> *vo sochne waale thhahare "andar ka hi kare **jaise** mein karti thi na kaam (**vaise)**"*
> She keeps thinking, "My daughter-in-law should do all the housework **the way** that I did", *na*?
> **Fig 131: Urmila (F47), convo_g**

The utterance is a quasi-quotative, and the literal meaning 'in the way in which I did *na* the work' would technically be concluded with correlative *vaise* 'in that way' to complete the construction. However, the correlative is often dropped in informal dialogue and the meaning still fully understood.

8.5   PART IV Functions of *type*: hedge

8.5.1   Hedging in Hindi

As with many languages, Hindi has a variety of hedging strategies. Lexical hedges include the adjective and adverb *thoda* 'little; a bit, a little' and the quantifier *kuch* 'some'. When followed by the hedge *sa*, as in *thoda sa* (N=82), the adjective functions unequivocally as a hedge, but both lexemes have hedging functions on their own as well.[93] Since *thoda* and *kuch* are not the focus of study in this thesis, the various functions of each lexeme have not been disambiguated, and the total token counts include 'original' uses as well as grammaticalized hedges (*thoda* N=460; *kuch* N=708). The following examples show *thoda* in different syntactic contexts, functioning as a hedge in all of them. In Fig 132 *thoda* appears in a canonical position modifying an adjective, but as part of a vague expression.

> ***thoda alag hi cheej*** *thi matlab gaon ki*
> It was **a bit of a different thing**, I mean, about the village
> **Fig 132: Anita (F32), convo_b**

The discourse context of Fig 133 is the reporting of a conversation with a neighbour, Dolly (pseudonym), whose husband was waking her up every night at 2am to keep him company while he went on duty as a trainee soldier. Both recording participants felt sorry for Dolly, and the use of *thoda* in the telling of the story by Ankita sheds doubt on whether she believes Dolly's husband will really stop calling her in the middle of the night once his training period is complete.

> *do baje se duty hai uski, phir-*
> His shift starts at two o'clock [in the morning]. Then-
>
> *ab **thoda-** jab training puri ho jayegi, phir to nahi hoga uska utna kaam*
> Now **kind of-** When his training is complete, then he won't have so much work
> **Fig 133: Ankita (F28), convo_d**

The position of the hedge at the beginning of the utterance ('Now, kind of-') hedges the whole utterance. Similarly, the use of *thoda* at the beginning of the sentence in Fig 133 hedges the entire sentence. The fact that *thoda* remains uninflected despite being juxtaposed with a plural noun is an indication of its grammaticalized status.

> ***thoda do chaar kamre*** *honge*
> There must be **like only a few rooms** (lit. little two-four rooms)
> **Fig 134: Vimla (F31), convo_d**

---

93 Token counts were retrieved from ELAN using regular expression (regex) searches such as these: '\bthod[a|e|i]\b' and '\bthod[a|e|i] s[a|e|i]\b'.

The example in Fig 134 illustrates that *thoda* can function as a hedge on quantity, a function not fulfilled by *jaisa* or indeed *type.* Other strategies for achieving quantity approximation in Hindi involve the use of adverbs such as *lagbhag* 'approximately' (N=14) *karib* 'almost' (N=11) and *kai* 'several', as in *kai log* 'many people' and *kai baar* 'many times' (N=32). The quantifier *kuch* 'some' (N=708), mentioned above, was not frequently used alongside numbers, but there were a few cases of *kuch* fulfilling a rounder function, such as in Fig 135:

> *jaise vo bata raha tha ki pichhle paanch chhe saalon me* **kuch** *pentis nahi das*
> *saalon me* **pentis officer mare hain** *CRPF ke aur is tarah ke forces ke*
> For instance, he was saying that in the last five or six years, **something like** thirty-
> five- no, in [the last] ten years, **thirty-five officers had died**, from CRPF and from
> those kinds of [armed] forces
> **Fig 135: Suraj (M25), convo_m**

However, it seems that in Pithoragarh, the primary strategy employed for quantity approximation is not a lexical hedge but the use of a fixed vague expression such as *do-chaar,* lit. 'two-four', which can refer to any number from two to eight, or the cultural norm of juxtaposing two numbers to indicate approximation, as in *pachaas-saath log the* 'There were about 50 to 60 people'. Vague communication is an important part of the culture, evidenced by the extensive use of phoric similatives with no specific discourse reference, and by the propensity to approximate numerically.

Jaisa as a hedge

Relative adjective and pronoun *jaisa* has been seen to function both as a phoric modifier and a similative, for which it acquires postpositional status. It has also developed extensive uses as a hedge for approximation and mitigation. In the following example, we observe both similative and hedging uses of *jaisa* as Bhuwan describes physical symptoms of a long-term tiredness he has been experiencing. The utterance has been split into two lines in Fig 136. In the first part of the utterance, we see Bhuwan unable to find the right word to describe the tiredness he had been feeling, which we understand from the context is unusual and chronic and has coincided with loss of appetite and other symptoms. Using *jaisa* as a hedge for approximation allows Bhuwan to communicate that the word *thakaan* 'tiredness' did not convey exactly what he meant. The example illustrates the use of *jaisa* without a following noun ("N + JAISA"). The corpus also contains an instance of the hedge *thakaan jaisi cheez* 'something like tiredness' with the pattern "N + JAISA + N"; however, this pattern is infrequently used with a hedging sense.

> *nahi vo ekdam **thakaan jaisi** sharir me*
> No, it [was] **kind of an extreme tiredness** in the body.
>
> *bilkul **aisa** ho raha tha ki pura sharir **tuta hua jaisa** na*
> It was totally **like** my whole body **was kind of broken**, you know.
> (lit. totally **like-this** was being that whole body **broken like**, no)
> **Fig 136: Bhuwan (M37), convo_b**

The second part of the utterance is a relative-corelative construction with *aisa* in which Bhuwan uses the metaphor *tuta hua* 'broken' to describe his body. The use of *jaisa* for incorporating metaphor into discourse is one of its key uses; this could be analysed as a similative or hedge and is perhaps a bridge between the two.

The utterance in Fig 137 is an example of the hedging use of *jaisa* to mitigate a negative description of another person. The speaker does not want to describe the woman as outright 'stupid' and hedges the insult with *jaisa*. The hedge communicates that the description is less than completely accurate, or that the speaker is not completely committed to the description (or both), reducing the effect of the word 'stupid' in a similar way to a hedge for approximation but for interpersonal reasons.

hin	वो	बेवक़ूफ़	**जैसी**	औरत	होती	होगी
tra	vo	bewakoof	**jaisi**	aurat	hoti	hogi
gls	She	stupid	**kind of.F**	woman.F	be.PRES.FSG	COND.FSG
eng	She must be kind of a stupid woman.					

Fig 137: Anita (F32), convo_a

As a similative *jaisa* usually (if not always) follows a nominal, whereas as a hedge, there are almost equal numbers of *jaisa* colligating with an adjective, as in Fig 137. This appearance following an adjective demonstrates incontrovertibly that *jaisa* is no longer a postposition and is evidence of grammaticalization not yet reflected in the lexical entry for *jaisa* in the OHED. However, it is not immediately clear which lexical category should be ascribed to *jaisa* in this context; it cannot be an adverb since adverbs always precede adjectives in Hindi. The working theory adopted here is that *jaisa* takes on a suffixal role similar to *-sa*, a morpheme which developed from *jaisa* and which may have had an influence on its colligation with adjectives.

Sa as a hedge

The suffix *-sa* evolved from *jaisa* as a similative, and a marker of intensification and approximation, and is syntactically highly flexible, although it shows a strong preference for

colligation with adjectives. The lexical entry from the OHED (Fig 138) lists the hedging usage ('-ish') as the primary sense of *sa.*

सा *-sa* **adj.**

resembling, like

1. rather, quite, -ish (moderate degree of a quality)
2. (chiefly w. adjectives of quantity, to indicate that an exact amount is not intended.)
3. that which; the kind of... which. - ~ कि, **conj.** just as; as

Fig 138: Lexical entry for *sa* in the OHED

Montaut (1995) describes the overall function of *sa* as follows:

1. *sa* transforms concepts that are not scalar but categorical, like 'to stop', 'dog' into scalar concepts

2. The most dominant characteristic of its uses is the 'affective' tone (expressive, emotional, etc) that it superimposes onto its semantic value of approximation. It is as much an expressive marker as it is an approximative

3. *sa* fulfils the same function as reduplication when used with colours and flavours – the sense of 'kind of' / approximately

4. The suffix *sa* has an additional sense when added to dimensional adjectives, quantifiers and positive adjectives like 'good' - that of **intensification**, e.g. *chhota-sa* = very small, *bada-sa* = very big. (This is reminiscent of the intensifying use of adjective 'quite' in English.)

As purely a marker of intensification (point 4), *sa* is not a direct equivalent to *type*, such as in the utterances in Fig 139 below where in both cases *sa* intensifies the adjective *chhota* 'small', both marked feminine to accord with the gender of the nouns 'problem' and 'girl'. The examples illustrate the expressive use of word order in Hindi: in (a) the intensified adjective *chhoti si* 'tiny problem' is fronted which brings it into focus; in (b) the principal noun 'girl' is focused, and the description of the girl as 'tiny' appears in an utterance-final tag, a common position for providing additional descriptive information.

a. ***chhoti si problem*** *agar ho gayi to...*
   If a **tiny problem** occurs...
b. *bacchi isi ki hai vo,* ***chhoti si****?*
   Is that **tiny little** girl his? (lit. girl his that **tiny little**)

Fig 139

However, lexemes are multifunctional and can fulfil multiple pragmatic roles at a time. There are several instances in the data in which *sa* when applied to dimensional adjectives functions additionally as a hedge. Tokens in which *sa* fulfils a purely intensifier role have not been included in the data used for comparison with *type*, but any sentence in which *sa* additionally fulfils a hedging function have been included. Fig 140 is an example of *sa* functioning as a hedge for approximation.

> *ek haar jeet ka ek **khel sa** ban gaya na*
> It's become a winning and losing **kind of a game**, hasn't it?
> **Fig 140: Suraj (M25), convo_m**

8.5.2 *Type* as a hedge

Hedging is one of the most prolific uses of *type* in the corpus, accounting for nearly half of the tokens (32/67). These can be divided into hedging for approximation (N=25) and hedging for mitigation (N=7), reflecting the varied motivation for employing the hedge. *Type* is not attested as a hedge on quantity, a usage referred to as a 'rounder'. One possible reason for this is that *type* has not yet reached that stage of grammaticalization, since rounder usage develops after other kinds of hedging (see the detailed overview in Chapter 7.2). Another possible reason was presented in the previous section (8.5.1): it seems Pithoragarh speakers have other strategies for approximating numbers and tend not to use a hedge.

Hedging for approximation

As discussed previously, hedging is a broad topic and even within the context of hedging for approximation there are many ways in which a hedge can affect the reading of an utterance and many hedging strategies which a speaker can employ. Several of these strategies are attested among the 25 tokens found in the spoken data from Pithoragarh: hedging for lexical approximation (not exactly the right word); flagging foreign or uncommon words; and hedging to distance oneself from what one is communicating.

Lexical approximation - not exactly the right word

There are a number of reasons why the word chosen by the speaker may not be (in their own estimation) exactly the right word to describe the item (or situation) in question. One reason might be that the item is not a prototypical example of the word selected. Another reason might be that the word does not 100% accurately reflect the situation but describes it in an approximate manner.

The pragmatic use of the *type* construction as a hedging device is evident from convo_k in which the recording between Tarun and his wife Rashmi is temporarily interrupted by a Nepalese labourer who is working in their back garden. The labourer, Prem, asks if he could keep some of the wood which he is clearing out of the garden, and there ensues a clarifying discussion between the two men as to which wood Tarun wants to keep and which is going spare and might be useful to someone else (Fig 141).

hin	हाँ	कोई	अपने	ऐम	झोंपड़ा	**टाइप**	का	बनाने	के लिये	ले	जा	सकता
tra	han	koi	apne	erm	jhopda	**type**	ka	banane	ke liye	le	ja	sakta
gls	yes	sb	REFL	FILL	hut	**type**	GEN	make	for	take	go	can
eng	Yes, someone could take it to make their, erm, hut kind of thing											

Fig 141: Tarun (M59), convo_k

Tarun's use of the *type* construction with *jhopda* 'hut' – a word for the permanent residence of many lower income families in India - indicates that the spare wood in his yard isn't quite the right sort of wood to make a proper *jhopda,* exactly the way they're normally made, but could be useful for someone to build something similar to a *jhopda*. His next utterance (Fig 142) is also in the present conditional mood ('could be useful') and reiterates the speaker's belief that the wood could be useful. Here, rather than repeating the *type* construction, Tarun uses another strategy for communicating vagueness in Hindi – echo compounding. His use of *jhopde-hopde* 'hut or a similar thing' rather than simply *jhopde* 'hut' is further indication that he is aware the wood will not make a prototypical hut, but he is nonetheless assured that it will be useful for making some kind of residence similar to a hut.

> *kisi ke jhopde-hopde me lagane ke liye bhi kaam aa sakta ye. kyon? hai na?*
> This could be useful for someone to put in their hut or a similar thing. Don't you think? Couldn't it?
> **Fig 142: Tarun (M59), convo_k**

Hedge and flag for foreign or uncommon words

One reason a speaker may be uncertain about the word they have selected (and therefore want to employ a hedge) is if the word is new to them or their interlocutor, or if they are in some way unfamiliar with its meaning(s). This may be the case with English-origin words which have not yet become established as part of the Hindi lexicon and could therefore be considered as (or closer to) code-switches on the code-switching-borrowing continuum. Borrowings which are well-established in Hindi such as *school* and *teacher* always appear without a hedge in the corpus (133 and 67 tokens respectively). In addition to well-established borrowings, there are hundreds of lexemes which could be considered newer

borrowings or code-switches and which appear much less frequently in the corpus. Many of these appear without a hedge, unmarked, but some are hedged with one of the Hindi lexemes mentioned in the previous section, and some are marked with *type*. Words which appear as part of a *type* construction and could be considered newer borrowings or code-switches include the following: nouns *attendance, celebrity, rock band,* and *value*; adjectives *cool, manly* and *sleek*; and verb *dominate*. Each of these words appears once in the corpus and is followed by *type* and its genitive clitic.

A discussion with one of the participants, Bhagwan from convo_e, throws some light on this pattern.[94] Bhagwan described his tendency to use a *type* hedge after some English words in this way: "I feel almost like a compulsion because it's a foreign word." As we discussed this further, he went on to describe this impulse as rooted in a fear of not using the word completely accurately: "In case I might not know the meanings of the word so well, I may not have the complete command of the word." He concluded that it was "safer to hedge" than to use a word in an incorrect or inaccurate way.[95] One sentence we discussed in which this may apply is the following, spoken by Bhagwan's fellow participant Manas (Fig 143). Since *sleek* is not a commonly used English word, it may be that Manas is not confident that it applies in this context. This hedge for English code-switches is therefore similar to the hedge for approximation.

> **back design** *bahut achha hai,* **sleek type ka**
> 'The design of the back is very nice, **kind of sleek**'.
> **Fig 143: Manas (M17), convo_e**

Bhagwan's experience of something "like a compulsion" to hedge English words is also pertinent from the perspective of conventionalized social norms and raises a further possible explanation for the co-occurrence of *type* with less-common English-origin words. The data tentatively suggests that they are <u>flagged</u> as 'foreign' by adding *type* (as a marker immediately afterwards. Lexical flags are attested in other languages including Italian (Rosignoli, 2011) and also Jersey Normal French where Jones (2005) reports that code-switches are more likely to be flagged than borrowings. The use of loanword *type* as the choice of hedge supports the view propounded by Backus (2005), that established loanwords somehow retain an element of foreignness even though they are fully integrated into the

---

[94] The discussion took place during a phone call in February 2019. All quotes were transcribed during the call and are reproduced word for word.
[95] Whilst Bhagwan's comments show great self-awareness, what must be noted is that he uses many English-origin words and hedges very few of them. An explanation for this may be that in most cases, as a skilled English speaker, he is confident as to the meaning(s) of the word.

Hindi lexicon. Speakers may be aware of this only at a sub-conscious level, and it is only through observing the 'behaviour' of a word through collocational patterns that we can conclude that it is somehow marked as foreign. One such pattern is the use of *type* as flag for foreignness. This hypothesis will be further investigated in Chapter 9, where the Hindi hedge *jaisa* is seen to accompany Kumaoni words.

Distancing oneself from what one is saying

Another motivation for using a lexical approximator is to create distance between her/himself and what s/he is saying. The use of a hedge indicates to the hearer that the speaker is unsure or even not in agreement with the content of their own communication, whether a single word or an utterance. In Fig 144, the three classmates from the higher-level English class at the ILC are discussing a recent visit to the institute by Bollywood actor Hemant Pandey, born in Pithoragarh. The speakers are discussing how ordinary the actor appeared and Chandrashekar and Manas are clearly surprised at this, given the unusual characters he has played; references to a strange noise, "blblblblbl", come from his comedic role in the hit film Krrish (from 2006). In lines #6 and #7, Manas continues to express doubt over the actor's fame, since he had not been approached by local fans. Later in the dialogue, the participants discuss whether he would be more likely to be asked for autographs in Delhi, in Pithoragarh or in one of the remote villages in the surrounding area. Manas' use of *type* in line #6 to hedge the adjective *celebrity* conveys his doubt about whether Hemant Pandey can truly be called a celebrity in the prototypical sense of the word, with his further comments focusing on the lack of paparazzi-style attention afforded the actor.

#	Speaker	Utterance
1	Chandrashekar	*aam aadmi lag raha hai vo*
	en	He seems like a normal man
2	Manas	*aam aadmi lag raha vo koi **common-***
	en	He seems like a normal man, some(one) **common-**[96]
3	Manas	*hemant pandey to us me dikha tha blblblbl waala aadmi*[97]
	en	But Hemant Pandey looked like a blblblbl kind of guy in that [film]
4	Bhagwan	*to tum kya chaahate ho ki jab vo ghar aaye tab blblblblbl bolte tabhi blblblbl karte hue aaye ya **dialogues** bolte hue aaye ya vaisa hi **dressup** karte hue aaye*
	en	So what do you want? That whenever he comes home, he comes saying "blblblblbl"? Or comes speaking **[film] dialogues**? Or comes in that kind of costume?

[96] This is not a grammatical clause in Hindi as it is in English. The intonation also suggests that a word is expected after common (a noun, according to Hindi syntax)

[97] The sound transcribed as "blblblblbl" is the sound made by moving one's finger up and down against one's lips while blowing raspberries, which was the movement produced by Hemant Pandey's character in the film Krrish.

5	Bhagwan	*asli me to sabko (hic) ek aam aadmi hi hai*
	en	In reality, everyone is just a normal guy
6	Manas	*aur agar vo- agar hota vo **celebrity type means** bahut **picturon** me aa chuka ho*
	en	And if he- If he was **sort of a celebrity, I mean,** if he'd been in a lot of **films**
7	Manas	*hota to koi to aata us se milne*
	en	If he was then someone would come to meet him
8	Bhagwan	*koi kya aata uske sath*
	en	Who would come with him?
9	Bhagwan	*aur vo already yahan aa chuka hai*
	en	And he's already been here before[98]

Fig 144: All participants, convo_e; all English-origin words highlighted in bold

The hedge both distances Manas from the word *celebrity* (a title ascribed the actor by others) and shows that he believes it to be less than completely accurate. This use of *type* as a hedge is reminiscent of the function of Russian *tipa* as a falsity marker, although without the 'pretending or deceiving overtones' which *tipa* can bring (Kolyaseva, 2021b:10). Note also that although *celebrity* is an English-origin word and the use of *type* is arguably also a flag indicating its foreignness and unfamiliarity, *celebrity* was already established in the discourse context due to previous mentions by each of the other recording participants and did not need to be flagged. This further supports the idea that the primary function of *type* was as a hedge for distancing and approximation.

The phrase *celebrity type* in Fig 144 is an example of elision (phonetic reduction), where the genitive postposition is removed from the end of the *type* phrase. Elision is one of the characteristics of grammaticalization and *type* constructions in other languages also show elision of adpositions as they grammaticalize. We have already seen how the removal of the category noun from the categorizer construction has led to postpositional and adverbial forms of *type* functioning as similatives and hedges. It is not surprising therefore that innovative uses of *type* without the genitive would also begin to appear.

There are four instances of 'bare' *type* in the Pithoragarh data including one instance in a stretch of monolingual Kumaoni dialogue (Fig 145). In all cases, *type* fulfils a hedging function.

---

[98] The addition of the English code-switch *already* duplicates the meaning of *aa chuka hai* which contains the completive modal verb.

kum	अजीब	**टाइप**	की	न	थी	ऊ	दोषिया	**पैचीला**	**जैस**
tra	ajeeb	**type**	ki	na	thi	oo	doshiya	**paicheela**	**jais**
gls	strange	**type**	GEN	NEG	COP.PST	3SG.PRO	rude	**irritable**	**kind of**
hin	ajeeb type ki nahi thi vo. *doshiya*, ***paicheela* jaisi**								
eng	She was **kind of strange**, wasn't she? Rude, **kind of irritable**								

Fig 145: Himani (F32), convo_i-2

The pronunciation of *type* in Kumaoni in Fig 145 was different to how it is usually pronounced in Hindi as there was an additional vocalisation on the word-final प /p/. This phonological integration suggests it may be a loanword in Kumaoni as well as Hindi, but more data is needed in order to confirm or deny this hypothesis.

Hedging for mitigation

Employing a hedge to modify the illocutionary force of one's utterance is a well-known communication strategy. Of the 67 *type* tokens in the corpus, seven of them are adverbial constructions that employ *type* to reduce the negative impact of the utterance. In all cases these are descriptions of another person, either pertaining to the person's physical appearance, or to their character or behaviour.

Hedging as a politeness strategy

Brown and Levinson (1987) pioneered research into hedging as a form of politeness, demonstrating through examples from Mayan language Tzeltal and from English how speakers use linguistic hedges to reduce the illocutionary force of a speech act, thereby mitigating any possible negative effects of their utterance. In the example that follows, two sisters-in-law are discussing where to buy leggings for the approaching monsoon season and Vimla comments on some leggings which Ankita had bought for her previously. The sisters-in-law live together and have a good relationship, and so do not need to be extremely polite with one another. Despite this closeness, a line-by-line analysis of the conversation (in Fig 146) reveals the manifold ways in which Vimla shows consideration for her sister-in-law's feelings by the way she structures and introduces her negative comment about the black leggings, and in her use of a *type* hedge meaning 'kind of'.

#	Speaker	Utterance
1	Vimla	*aaraamdaayak to <u>hain</u> hi ye waali, han? leggiyan jo tu laayi hai, ye waali*
	en	They <u>are</u> comfortable, these ones, you know. The leggings that you bought - these ones[99]
2	Ankita	*hmm*
	en	hmm
3	Vimla	*lekin ye waali kaali waali bahut **chust type ki** ho gayi hai*
	en	But… these ones[100], the black ones, have become really **kind of tight-fitting**
4	Ankita	*achhi nahi ho rahi hai*
	en	They aren't good

Fig 146: Ankita (28) and Vimla (31), convo_d

In Line #1, Vimla wants to reassure her sister-in-law that the pairs of leggings she had sourced were comfortable and she employs several pragmatic devices to this end. Lexico-syntactically, this reassurance takes the form of expressive word order which fronts (and focuses) the positive adjective 'comfortable', use of the emphatic particle *hi,* and use of the filler *han* 'yes', which functions as a question tag fostering agreement with the listener. Prosodically, this reassurance takes the form of stress on *hain hi* 'are' (emphatic), and a raised intonation on the tag *han*. Prosody also plays an important role in expressing Vimla's tentativeness in making a negative comment about the black leggings, since the conjunction *lekin* 'but' which introduces Line #3 is slightly elongated, showing some hesitation. The *type* construction acts as a hedge on *bahut chust* 'really tight-fitting' and plays a final part in mitigating any perceived criticism in her admission that the black leggings (in particular) were not very suitable. Vimla's politeness strategy seems to have been successful, as Ankita agrees with her that the black leggings are not good (Line #4), and the pair turn to discussing which shops are good for buying leggings.

Syntactically, hedge *type ki* modifies the adjective *chust* 'tight-fitting' (indicated by the blue box in Fig 147). The genitive appears in its feminine form *ki* agreeing with the gender of loanword *leggi* 'leggings', the subject of the sentence, which had been specified in the previous utterance and is represented by the phoric phrase *ye waali* 'this one'. Hindi adverbs ending in *-a* are marked for gender (including hedge *jaisa*), so this lends support to the interpretation of *type ka* as an adverbial hedge agreeing with the sentence subject.

---

[99] In Pithoragarh dialect, the verb *lana* 'to bring' is also used to mean 'to buy', possibly due to elision of the full phrase *kharidke lana* 'to buy and bring (home)'.

[100] In Hindi, the loanword *leggi* '(pair of) leggings' is a singular noun, so Line 1 shows the plural form of the noun whereas in Line 3 the speaker is referring to only one of the pairs of leggings that were bought.

hin	यह	वाली	काली	वाली	बहुत	चुस्त	टाइप	की	हो	गयी	है
tra	ye	waali	kaali	waali	bahut	chust	**type**	ki	ho	gayi	hai
gls	this	one	black	one	very	tight-fitting	**type**	GEN	become	INTF. PST	AUX
eng	This one [pair of leggings], the black one, has become really kind of tight-fitting										

Fig 147: Vimla (F31), convo_d

The speaker's use of the booster *bahut* 'really, very' with adjective *chust* 'tight-fitting' reveals the true state of the leggings. The *type* hedge is clearly not an example of hedging for approximation such as we encountered in the previous section, since the speaker is very sure that the leggings are too *chust* 'tight-fitting'. Rather, the hedge affects the speech act of 'giving negative feedback about something someone bought for me' by downplaying the negative aspect of the item purchased and minimising its impact. This example illustrates the importance of analysing the utterance in context in order to correctly interpret the pragmatic intent of the speaker and the effect created by the *type* construction.

Mitigating negative descriptions of people

In Fig 148, three neighbours are discussing the (newly orphaned) teenage children of a neighbour who had died that week, a sister and brother. We join the conversation when they are discussing the sister, and they quickly move on to talk about the brother, Deepak:[101]

#	Speaker	Utterance
1	Preity	*lag raha hai usko dekhke kuch bhi nahi lag raha hai yaar*
	en	Did it seem like it from looking at her? It didn't seem like it at all, *yaar*.
2	Garima	*yahi se dekha meine bacche hi ajeeb jaise hain yaar vo deepak bhi to vaisa hi*
	en	The children themselves are kind of strange, *yaar*. That Deepak is also exactly the same
3	Anita	*adiyaat* **type** *ka*
	en	**kind of** idiotic
4	Preity	*hain dono?*
	en	Both of them? Really?
5	Garima	*kahan lag raha tha phir [deepak] ko bhi…*
	en	It didn't seem like [Deepak] had just lost his mum either…

Fig 148: All participants, convo_d

The construction with *type* appears in Line #3, in which Anita interjects in support of Garima's assertion (in Line #2) that the children are both 'kind of strange', describing the boy (Deepak) as *adiyaat type ka* 'kind of idiotic'. When I asked Anita later why she had described the boy in this way, she explained that Deepak wasn't at all social and would never speak to anyone. However, she didn't want to say that he was <u>definitely</u> an idiot; that would be too strong a

---

[101] Pseudonym

statement, and she couldn't be so sure that it was justified. Her comments show how the *type* hedge communicates on multiple levels simultaneously: Anita follows what seems to be social politeness protocol in hedging negative descriptions of other people and employing a hedge construction with *type* allows Anita to express a subjective opinion without having to be definitive or categorical about it.

Hedges not needed for positive comments

We have established that negative comments about another person are often mitigated by a hedging strategy. However, in order to be certain that the hedges are used for mitigation of negativity, instances of positive comments about other people have also been analysed. We can observe that these are expressed directly, without recourse to a hedge.

The most frequently used affirmatory adjective in Hindi is *achha* which means 'good', but also 'ok', 'right' and 'nice', and has furthermore grammaticalized into a conversational filler and an expression of surprise similar to 'oh really?' A detailed search of the corpus reveals 1540 tokens of *achha* including all of its possible inflections, among which there are no examples of *achha* being followed by the hedges *type* or *sa* and only one example of *achha* being followed by *jaisa* (Fig 149). In this conversation, the speaker is describing a situation in which her sister had turned up at a party wearing leggings that were too tight and looked very odd. On her advice/instruction, her sister had subsequently changed into a looser pair of leggings and then looked *achhi jaisi* 'kind of nice'. The use of the hedge indicates that the sister did not look amazing but looked much better than before. In the context of the narrative, therefore, the use of a hedge with *achha* 'good, nice' to describe another person is affirmative and not derogatory. The speaker's prosody supports this interpretation: *achhi jaisi* 'kind of nice' was spoken at a slightly higher pitch expressing an element of surprise.

#	Speaker	Utterance
1	Vimla	*meine kaha, "kholo ranjani. ja tu dhila waala pahan"*
	en	I said, "Take them off, Ranjani. Go and put on a looser [pair]."
2	Vimla	*phir dhila waala pahana. phir **achhi jaisi** lag rahi thi*
	en	Then she put on a looser [pair]. Then she looked **kind of nice**.

Fig 149: Vimla (F31), convo_d

In addition to a search for examples of hedging in collocation with *achha* 'good, nice', all tokens of the adjective *sundar* 'beautiful, handsome, attractive' were also extracted from the data. The adjective can be used to describe the physical appearance of both males and females as well as views and paintings, etc. The table in Fig 150 indicates the general subject

matter of the referent described by *sundar*, of which the most common is another person's physical appearance and specifies how many tokens of the adjective were modified by a hedge or a booster, or indeed neither. Analysis shows that senses of the word are universally positive, and in line with expectations, of the 29 tokens in the corpus, not a single usage is hedged. Rather, 20/29 instances of the word appear with a linguistic booster and the remaining nine tokens are of the unmodified adjective.

Referent	# Tokens Booster	# Tokens Hedge	# Tokens Neither	Total
Another person - physical appearance	9	0	5	14
Clothing	6	0	4	10
Domestic animal	2	0		2
House	1	0		1
Picture of idol	1	0		1
Temple	1	0		1
Total	**20**	**0**	**9**	**29**

Fig 150: Boosters and hedges as collocants of *sundar* 'beautiful'

Boosters employed by speakers consist of 11 instances of the adverb *bahut* 'really, very' or alternatives *bada* 'big' and *ati* 'very much', as well as 9 instances of *itna* 'so much' and its interrogative partner *kitna* 'how (much)', used similarly to English 'how beautiful'. This brief analysis of positive adjectives illustrates the difference between the lack of hedging that accompanies them compared with the use of hedges with negative adjectives.

Gender differences in hedging

A brief analysis of gender differences in the data showed that women hedge more than men. The use of hedging devices as a politeness strategy is well documented (Brown and Levinson, 1987) and more recent work has shown that women, expected to be more polite and less direct in their language use, achieve this politeness through a variety of code-switching practices (Gardner-Chloros, 2009:84-87). However, considering the different kinds of hedging separately, we find that men and women showed an identical number of tokens for hedging for approximation purposes, whereas all tokens of hedging for mitigation were associated with women. Furthermore, these findings are conflated with choice of conversation topic: a large proportion of the hedges for mitigation were used when describing another person's appearance, a topic wholly preferred by women. Since males and females tend to discuss different topics, the need to hedge for mitigating purposes is dictated by choice of topic, hence the gender bias in the numbers. Additionally, analysis of

individual speakers revealed idiolectal differences, with female speakers Vimla and Ankita between them accounting for nearly 60% of all hedging for mitigation tokens.

At the same time as considering the hypothesis that gender plays a part in the employment of hedges (and therefore in the ongoing use of *type* as a linguistic hedge), we must also consider the possibility that the predilection for using vague language is a feature of Kumaoni society. Indeed, there is some evidence that dialectal differences may be a factor. Results of the acceptability survey I carried out amongst university age students in Pithoragarh show that the phrase *pagli jaisi* 'kind of crazy' is very commonly used. By contrast, Gaura, a research informant from another state in North India, revealed that in her city (Varanasi) such a phrase would rarely be used. She explained that in Varanasi, people are more direct in their manner of speaking; rather than hedging the insult, they would just say, *Are paagal hai!* 'Oh, she's crazy! Cross-dialectal studies would add to our understanding of how *type* and other hedges are used across North India.

8.6 PART V Incipient functions of *type*: quotative, general extender and filler

8.6.1 General extenders in Hindi

There are several strategies for extending discourse text in Hindi. Reminiscent of how numerical approximation can be achieved without recourse to a hedge by juxtaposing numbers, as in *do chaar din* 'several days' (lit. two four days), reduplication is one way of indicating that the item specified is approximate and should be understood as representative of a set of similar items. Common reduplicated words in colloquial Hindi include *kapade-(h)apade* 'clothes and things'*, bartan-vartan* 'pots and pans and things' and *sabji-(h)abji* 'vegetables and things'. *kapade-(h)apade* is usually attested in the context of a narrative in a phrase like *meine kapade-(h)apade dhoye* 'I washed the clothes and things' and could easily include bedsheets and curtains; whatever items were being washed along with the clothes. In another context, such as packing for a holiday, a phrase like *meine kapade-(h)apade pack kiye* 'I packed my clothes and things' could conjure up other items such as a toothbrush and a phone charger.

The most common general extender attested in the recorded data from Pithoragarh is the centuries old Perso-Arabic loanword वगैरा *vagaira* 'and the rest, and so on'. *Vagaira* appeared 43 times in the corpus, in each case following a single noun and denoting a set of items similar to or related with the specified noun, as in Fig 159 where *vagaira* denotes snacks similar to popcorn which could be eaten after watching a film.

> *phir dekhke hamne **khub popcorn vagaira khaaye** chaai pi*
> Then having watched [the film] we **ate loads of popcorn and things**, drank tea
> **Fig 159: Usha (F59), convo_q**

There is also some evidence that *vagaira* is grammaticalizing and used where a category or set cannot be retrieved from the discourse, such as in *meine khaana vagaira khaaya* 'I ate food and things', where it is implausible that the speaker (Bhuwan) ate anything other than food.

Vagaira seems to have a monopoly on general extender usage in Hindi, although there were a couple of tokens in the corpus where *aise* appears to fulfil a similar function or support the more metatextual general extender function as a topic closer, as in Fig 160 where it appears just after *vagaira*.[102]

---

[102] The Hindi equivalent of English etcetera is not used colloquially, although etcetera has started to appear in informal conversation recent years but was not attested in the corpus.

ab ye hai ki thoda sa shuru me jara sa saa- **vo jaisa** *ho raha tha jyaada bolne me chalne me* **saas vagaira** **aise**

Now it's like this: a little bit in the beginning, ever so slightly, breathin- became **like that**, if I spoke too much, walked too much, **my breath and things,** **like that**

Fig 160: Leela (F55), convo_a

8.6.2   *Type* as a general extender

One of the *type* tokens can be best analysed as a general extender, following the broader definition adopted in the previous chapter. *Type* appears as part of the manner adverbial phrase *is type se* 'in this way, like this', but has an extended meaning best translated as 'this/that kind of thing'. The phrase appears during a discussion with the researcher (myself) and Viren's friend and his family about a new grading system introduced that year by one of India's largest high school exam boards (Fig 161). Having been asked directly by one of the family about whether it was possible for students to fail under the new system, I explained that it was no longer possible to fail, but students could still get bad marks, which I started to list in line #4, *jaise D, E, F…* 'For example, D, E, F…' However, I had only got as far as 'D, E…' when Lokmanyu interrupted in agreement with the *type* phrase: 'E, that kind of thing…'.

#	Speaker	Utterance
1	Lokmanyu	*gande number laate hain ye fail nahi hote hain*
	en	They get bad marks. They don't fail!
2	Viren	*CBSE board me na? CBSE board me na?*
	en	In the CBSE board, right? In the CBSE board, right?
3	Lokmanyu	*han na!*
	en	Yeah!
4	Researcher	***jaise*** *D E F*
	en	**Such as** D, E, F
5	Lokmanyu	*E **is type se** han*
	en	E, **that kind of thing**, yeah

Fig 161: Lokmanyu (M19), Viren (M20) and the researcher, convo_h-3

On the one hand, the choice of postposition **se** instead of the genitive **ka** is curious, and it is possible that the speaker used *se* an alternative to *ka* with the same intended meaning and without any intention of creating an adverbial phrase. In fact, a true adverbial reading is impossible, since there is neither a verb or an adjective to modify in Lokmanyu's utterance. However, rather than a component-based interpretation of the *type* phrase in which *type* would still be analysed as a noun, and the adverbial requiring a verb to modify, *is type se* could be interpreted as a cliticized unit (see Fig 162), a grammaticalized similative demonstrative of manner like *aise.* It seems that Lokmanyu is connecting his utterance to the researcher's utterance through a general extender-style commentary on the grades. It

may be that the use of *se* and the creation of a similative demonstrative of manner phrase with *type* is a deliberate pairing with the researcher's use of the manner similative demonstrative *aise*.

hin	ई	इस	**टाइप**	से	हाँ
tra	E	is	**type**	se	han
gls	E	DEM	**type**	ADV	yes
eng	E, that kind of thing / like that, yeah				

Fig 162: Lokmanyu (M19), convo_h

The second case of general extender usage (Fig 163) also involves the phrase *is type se* and appears at the end of a discussion amongst two friends about why (they think) the researcher is interested in recording them speaking Hindi![103] Having given (insightful) examples of her own innovative language use, Monika makes a statement to which her friend is silent, and Monika herself concludes the conversation using a series of conversation-ending strategies, including the general extender *is type se* and a fixed expression with *aisa* which is used for the purpose of concluding a narrative or point.

> ***vaise*** *ek hi shabd ko alag-2 log alag-2 taarike se kahate hain*
> **In fact**, different people say the very same word in different ways.
> *[1.5 second pause]*
> *na,* ***is type se,*** ***aisi*** *hota hai*
> Don't they? **Like that**. It happens **just like that**.
> **Fig 163: Monika (F29), convo_f**

As with the previous example, the *type* phrase is a single unit in the same form as the manner similative demonstrative but cannot be analysed as such because of the distance from any discourse content that could be deictically retrievable. Rather, the usage of the phrase fits the broad definition of a general extender given in the previous chapter: a topic-closing device which adds 'an element of vagueness [to] the proposition' (Aijmer, 2002:13).

---

[103] The author was not present.

8.6.3 Quotatives in Hindi

The quotative system in Hindi, as in many languages, revolves around the verb '(to) say' (कहना *kahana*), which is used in all tenses including the simple past and the present continuous (or progressive) tenses. Other verbs used include 'tell' (बताना *bataana*) and 'ask' (पूछना *puchhna*).

> ***vo*** *pahale se* **<u>*kahane waale*</u> *hue* *naukri mat kar …*
> **He had been saying** from before, "Don't work" … (lit. 'he … <u>the saying one</u> became')
> **Fig 151: Meena (F59), convo_g**

The dialect of Hindi spoken in Pithoragarh (and surrounding area) also boasts a number of non-standard quotative forms which are associated with speakers who have a strong attachment to the local language, Kumaoni. Non-standard forms using Hindi lexemes include *kahane waale hue* (Fig 151) and the localized form *kah rahe thhahare* with the verb ठहरना *thhaharna* 'stay, remain' appended as an auxiliary in place of the standard copula होना *hona* 'be' (Fig 152). The underlined words are given in a literal translation as even an idiomatic translation in English cannot capture the essence of the speech style which is highly idiosyncratic.[104]

> *aap hi ko bata rahi hoon* **kah rahi <u>thhahari</u>**
> "I'm only telling you", **[she] said** (lit. '<u>remained</u> saying')
> **Fig 152: Ankita (F28), convo_d**

Another notable feature of colloquial discourse among speakers with a strong link to Kumaoni, is the extensive use of the Kumaoni quotative marker भल/बल *bhal/bal*, which is used to mark information received from an absent third party, including quoted speech, and can be interpreted as both a quotative and an evidentiality marker. Fig 153 illustrates its usage in Kumaoni during a section of code-switched text (<u>underlined</u>) from one of the recordings:

---

[104] A Google search provides evidence of the non-standard nature of these forms: कहने वाले हुए *kahane waale hue* resulted in 2,960 hits, whereas a search using the standard auxiliary कह रहे थे *kah rahe the* '(He) was saying' retrieved 1.9 million hits. There was only one single hit on the localized form कह रहे ठहरे *kah rahe thhahare* (using *thhaharna* 'remain' as an auxiliary), from a website located in the Kumaon region; the article also includes a very high number of tokens of the verb as an auxiliary alongside the verb करना *karna* 'do' as well as fulfilling the function of a main verb in the sentence: https://www.nainitalsamachar.org/nationalist-government-first-time/ Accessed on 21st February 2021. This illustrates how localized this quotative form is.

gusse se kaha unhone "ki kuthu pe ki kuthu" **bhal** *phir*

She said angrily, "<u>What should I have said, then? What should I have said?</u>" <u>**she said**</u>, then.

Fig 153: Ankita (F28), convo_d

In Hindi, as well as Kumaoni, loanword *bhal* always appears after the reported text and often punctuates speech, as in Fig 154 where it appears after each new piece of information. The excerpt in Fig 154 is taken from the same conversation as Fig 148; here the neighbours are discussing what would happen to the orphaned children of their deceased neighbour who apparently did not have good relationships with her extended family.

#	Speaker	Utterance
1	Anita	*tabhi to sab se mil ke rahana jaruri hota hai, dekha!*
	en	That's exactly why it's so important to live harmoniously (lit. together) with everyone. See!
2	Preity	*chaachi logon se bhi inka koi vohi nahi hai* **bhal** **na** *apne parivaar me, devraani jethhaani se, koi mail nahi hai* **bhal**
	en	She doesn't have any of that with her aunt's family either **they say**, **you know**, in her family, with her sister-in-laws, **they say**

Fig 154: Preity (F42), convo_b

The vague use of the third person singular with the emphatic particle *vohi* denotes 'good relationship'. The sentence as a whole is vague, since the identity of the third-party is not mentioned, nor perhaps even known, as in the example above. It is a feature of narrative in the region that where the identity of the source it is not important, it can be omitted. The 'absent third party' quotative and evidentiality marker *bhal* suffices.

With the exception of Kumaoni loanword भल *bhal*, which always appears after the quoted material (in both Kumaoni and Hindi), the positioning of other reporting verbs including new quotatives can appear either before or after the quotation, often both.

Finally, similative demonstrative *aisa* also functions as a quotative. *aisa* always appears before the reporting verb in its position as an adjectival or pronominal similative demonstrative, regardless of whether the reporting verb comes before or after the quote as the examples in Figs 155 and 156 illustrate. Note the duplicate use of reporting verb *kahana* both before and after the quoted material in Fig 156.

BEFORE QUOTED MATERIAL
to puchh rahi thi matlab "tuition kyon nahi aaye karke". aur **aisa** *bol rahi thi-*
So [she] was asking, I mean, "Why didn't they come to tuition?" And [she] was saying **something like that**-

Fig 155: Neelima (F16), convo_n

169

AFTER QUOTED MATERIAL

*meine **kaha** "na, mein **aise** kaam nahi karti hoon" **aise kaha***

I **said**, "Nah, I don't do **those kind of** things", I **said like**.

Fig 156: Urmila (F47), convo_g

Perhaps having developed from a marker of imitation, as has been suggested for other new quotatives, or as an approximation, or both, *aisa* and *aise* are both used to mark non-lexical exclamations and gestures or both. They also both mark internal thought and hypothetical speech, as in Fig 157, where tokens of *aisa* juxtapose the quoted text.

*tumhaari report karte hain **kah rahe hain** "ye hai so hai" **aisa waala** han*
*ki "vo nahi aate hain" **aisa***

"We're gonna report you", **she said**, "**And this and that**", **such kind of things**, yeah?
Like "She doesn't come [to work]", **like that**.

Fig 157: Ankita (F28), convo_d

The example in Fig 157 is a complex one, and an accurate reflection of the colloquial Hindi used in Pithoragarh, involving multiple quotative strategies. The complementizer *ki* introduces a dependent clauses, often a quotation, and has here been idiomatically translated as 'like', rather than literal 'that', which is not used in this context in English.

The text in green is an example of a 'placeholder' quotation, where determiners are used to represent (take the place of) the actual quote. Note that in other utterances *aisa* is used in conjunction with its distal counterpart *vaisa* in this way, leaving the hearer to extrapolate as to what may have been communicated, for example, in *vo "aisa aisa vaisa" likh rahi hai* 'She's writing "this and that and this and that", where the reduplication of *aisa* suggests many things were written. *Aisa* and *vaisa* are also used together as a placeholder for actions, as in the English *I did this and that*.

8.6.4 *Type* as an incipient quotative

The subjects of new quotatives and non-standard quotatives in Hindi have been unexplored to date and are worthy of a greater emphasis than can be afforded them here. The regional dialect of Hindi spoken in Pithoragarh and surrounding hills shows a predilection for innovative quotatives, and into this rich tapestry of quotative forms enters English loanword *type*. Although there are only two examples of incipient quotative usage in the corpus, the use of *type* in this way, especially amidst so many alternatives, is potentially significant. In both cases, the context is perceived attitude or hypothetical thought rather than speech, which suggests an incipient stage of grammaticalization following the pattern discussed in Chapter 7.2. The use of incipient quotative *be like* to mark thought is believed to be where its usage emerged (Buchstaller, 2014:110).

In one conversation, such an expression used to comment on imaginary internal reported speech of a third party. It is to this example we now turn. Halfway through recording convo_n, two teenagers start on a form of social commentary, complaining how older, more traditional women make 'a big deal' out of them having male friends, even to the point of pronouncing doom on their futures. The girls respond with a fictional rebuttal which has them employing different voices, including fake pity and mocking sarcasm. The woman's response is also anticipated and acted in a third, shocked voice, printed in Fig 158 below.

Akanksha	*unko bhi realize hona chahiye na,* **ki** *"hain?! kya bolke gayi hamse?" is type ka*
en	Yeah, she should realize [what's she done], shouldn't she? **Like**, 'What?! What have they said to me?' **That type of thing**[105]

Fig 158: Akanksha (F17), convo_n

The function of the *type* construction is to mark the quoted (imaginary) speech as an example of the 'type of thing' that the woman might think to herself. This standalone phrase is intentionally vague and seems to have become conventionalized as a calque on native Hindi noun *aisa* 'such a thing (as this)' itself inherently vague and frequently used at the end of an utterance. Both this and the other token of quotative *type* in the corpus use the proximate similative demonstrative construction *is type ka* 'of this type, like this', one token appearing before the quoted text in combination with complementizer *ki*, and the other appearing after the quotation. This highlights again the prolific nature of the similative demonstrative *type* construction observed in the data.

8.6.5 Focus markers and fillers in Hindi

Hindi has many discourse pragmatic markers and it will not be possible to discuss them all here. One very frequently used discourse marker is the Hindi noun *matlab* 'meaning' (N=826). Many of the uses of *matlab* show persistence of the 'meaning' sense of the noun from which the DM originated. The context of Fig 164 is a conversation about Leela's recent poor health; Archana wishes to check that her aunt is not experiencing the same problems as before. Having enquired in one way, Archana then uses *matlab* 'I mean' to reformulate her question.

1	Archana	*abhi to pareshaani nahi hai koi*

---

[105] My idiomatic translation uses the distal demonstrative in English, whereas in Hindi the proximal was used. Cross-linguistically, it is typically the proximal demonstrative which develops cataphoric functions, including uses as a quotative (König, 2017:159).

	en	But right now there aren't any issues (lit. problem not is any)
2	Archana	*matlab jo pahale thi vo ab nahi ho rahi*
	en	**I mean**, what was there before isn't happening now.
3	Leela	*ab nahi ho rahi*
	en	It isn't happening now.

Fig 164: Archana (F32) and Leela (F55), convo_a

As well as reformulation*, matlab* is used to clarify someone's meaning and restate one's own. Younger speakers also use *means*, perhaps understood to be the English translation of *matlab*, as a discourse pragmatic marker in the same contexts. There was only one token of *means* in the corpus (see Fig 144, line #6) which the speaker used for restating his point in more detail.

Matlab is also used as a filler to buy time, as in Fig 165 where later in the recording Leela questions her niece on her salary.

#	Speaker	Utterance
1	Leela	*vaise mein puchhna nahi chaahiye lekin kitna de rahe hai ye*
	en	**Like**, I shouldn't ask, but how much are they paying you?
2	Archana	*khaali hua vo to, timepass hua*
	en	It's not for any purpose, though, it's a hobby.
3	Anita	*timepass hai xx*
	en	when I spoke too much and walked around, my breath and stuff, like that.
4	Leela	*phir bhi- nahi nahi phir bhi* **matlab vaise** *puchh rahi hoon*
	en	No no, still, **I mean**, I'm asking **just like that**
5	Archana	*saadhe tin haazaar*
	en	Three and a half thousand [rupees per month]
6	Leela	*are theek to hai yaar!*
	en	Oh but that's ok, *yaar*!

Fig 165: Archana (F32) and Leela (F55), convo_a

The use of *matlab* by Leela allows her to restate her request and get the information she was after (her niece's salary). Aware that she was asking an uncomfortable question, she responds in a very friendly way addressing her niece as *yaar*, which is unusual for an aunt and expresses solidarity, softening the effect of having forced Archana to state her (meagre) salary.[106]

---

[106] The only other time Leela addressed her niece with *yaar* in the recording was when she gave her medical advice. It seems that using a friendship term of address closes the distance in status and age and makes the content of the communication easier to accept.

Another Hindi discourse pragmatic marker has developed from the highly polysemous intransitive verb चलना *chalna* 'go (move), walk, work (function)'. The second person informal form of the imperative *chalo!* is used to express agreement with a plan, as in *chalo, theek hai* 'go on, that's fine', but also to express that something has turned out alright in the end, or that the speaker has changed their views and now agrees with the opinions or (particularly) plans suggested by another. There were 140 tokens of *chalo* in the Pithoragarh data, of which very few had an imperative meaning, and the majority had an innovative discourse usage.

#	Speaker	Utterance
1	Rashmi	*shaam tak aaye<u>nge</u>*
	en	They'll come by the evening.
2	Tarun	*shaam tak. **chalo** theek hai*
	en	By the evening. **Alright then**, that's fine.

Fig 166: Rashmi (F53) and Tarun (M59), convo_k

In Fig 166, Tarun was concerned about the time of arrival of their expected guests and wanted to phone and check if they needed picking up, but his wife Rashmi persuaded him that they would come on their own and by the evening. She repeated her assertion no fewer than four times 'they'll come by the evening', and eventually Tarun agreed to let the point go. *Chalo* expresses this concession to his wife's view of the situation and her plan to wait for the guests to arrive.

Aisa as a focus marker and filler

It has been shown throughout the chapter that lexeme *aisa* is highly polyfunctional, as has been observed for Portuguese similative demonstrative *assim* (Bittencourt, 1999; Lopes and Carapinha, 2004), and that *type* seems to be developing corresponding functions. Although not prolific, *aisa* has been attested as a focus marker in the Pithoragarh data in cases such as the first token in Fig 167. (The latter token is an example of quotative usage.)

> *aur ab sab logon ko **aisa**, **you know**, <u>study</u> vahi lagta hai ki "han study ki field me hi aage jaye" **aisa** to*
> And now everyone, **like**, **you know**, feels just that about <u>studies</u> [that] "Yes, we should only go into the academic field", **like that**[107]
> **Fig 167: Akanksha (F17), convo_n**

---

[107] There is no direct English equivalent to the structure used in the Hindi *study vahi lagta hai ki "[QUOTE]"*. In the Hindi, *vahi* lit. 'that only' is anaphoric, and in this context specifically cataphoric, since it is followed by complementizer *ki*. The quotation (verbalized thoughts or attitude) introduced by *ki* provides the content for the verb *lagna* 'feel, think, seem'. The use of *ki* to introduce quotations is illustrated in Chapter 8.6.3.

The use of English code-switch *you know* as a supplementary discourse pragmatic marker clustering with *aisa* creates a larger pause before stating the main topic of the utterance *study* 'studies', which is thereby emphasized.[108]

Aisa does not commonly appear a filler, perhaps because of the monopoly held on the filler role by *matlab* and *vaise* and perhaps also because of the crucial role it plays in structuring discourse. It rarely appears empty of semantic content.

8.6.6   *Type* as a filler

The examples of incipient quotative marking in the Pithoragarh data suggest that the *type* phrase can be positioned either before or after the quoted material, but must be in close proximity to it, similar to Russian where it must be immediately adjacent (Kolyaseva, 2018). This seems logical: for the construction *is type ka* to fulfil its function as a quotative <u>marker</u>, especially given its highly polysemous nature, the phrase must be positioned close enough that the hearer is in no doubt as to what content is being marked. Similarly, to function as an utterance-closing general extender, although we are to expect a slight gap, the general extender needs to be spoken close enough to the rest of the utterance to be perceptively connected by the hearer.

Focus marker and filler functions are less restricted, and their syntactically extra-clausal (standalone) nature is an indication that further grammaticalization seems to have taken place. There is only one token of unambiguous filler usage in the Pithoragarh data. In Fig 168, a *type* phrase with the proximal demonstrative *is type ka* 'kind of' is seen to function as a filler, sandwiched between two other fillers.

> *matlab,* **is type ka,** *matlab*
> 'I mean, **that kind of thing / like (that)**, I mean'
> **Fig 168: Mohan (32), convo_j-2**

Deictically, *is* 'this' could possibly relate to the word 'dumb' which the speaker was trying to say, and which eludes him in the subsequent utterance. An argument is often given that discourse markers are semantically bleached of the lexical content of their source words; however, this is not universally the case. The persistence of (traces of) lexical meaning in grammaticalized forms of a word is a documented feature of grammaticalization and certainly not proof against it. Having established that the phrase *is type ka* has become its

---

[108] There were 34 tokens of *you know* in the recording, all by the same speaker (Akanksha, F17). The discourse marker is clearly part of Akanksha's idiolect, but not common in Pithoragarh; as it was only used by two other speakers in the corpus (N=1; N=2), both English-medium educated and both younger than 30.

own indivisible lexical unit with a meaning akin to 'like this' (see the section on phoric similatives, 8.3.5), we observe how elements of this semantic meaning may persist as the phrase is extended to additional pragmatic contexts. A longitudinal study in another 4-5 or 9-10 years (or both) would garner a greater number of discourse pragmatic tokens and allow us to analyse how the *type* phrases have developed or become established, if at all, and with what meaning(s).

8.6.7 Other semantic-pragmatic functions

In addition to the functions shared by *type,* there are innovative functions of *aisa* and related lexemes which are not shared by *type.* The primary focus of this thesis is to investigate uses of *type* in colloquial Hindi, and it is therefore not possible to dedicate much space to semantic-pragmatic functions that do not concern *type*. They will be mentioned briefly here:

- The **exemplifier** function ('like, such as') commonly developed by *type* cognates cross-linguistically is fulfilled by the relative form of the similative demonstrative of manner, *jaise*

- The distal form of the similative demonstrative of manner, *vaise*, has developed several **discourse-structuring functions** focused on opposition or the unexpected, i.e. to politely offer a new viewpoint

- The intensifier function of *aisa*, common among similative demonstratives, as in Fig 169:

  > *mere ko to **aisi** daaya aati hai*
  > I feel **so much** compassion [for her].
  > **Fig 169: Vimla (F31), convo_d**

- The discourse-structuring role of *aisa* and a summary of **key fixed expressions** in this domain are discussed in the following section

Discourse-structuring functions of aisa

Grammaticalized similative demonstrative *aisa* is used to form formulaic expressions functioning as discourse-structuring devices, many of them showing remarkable similarity to discourse-structuring uses of German similative demonstrative of manner *so* (König, 2020). In fact, 55 out of a total of 609 *aisa* tokens in the data appear in one of these idiomatic fixed expressions. The expressions (Fig 170) are divided into 'introductory, initiating' discourse-structuring functions (a-c) and 'responsive, conclusive' functions (d-h), following König (2020).

Introductory, initiating functions

a) *aisa hua (na); aisa hua ki …*
That's what happened, right; What happened was …
Starting a narrative or explanation about the past

b) *aisa ho raha (hai) (na)*
That's what's happening (isn't it?)!
Starting a narrative or explanation about the present

c) *aisa hai (na); aisa (nahi) hai ki …*
It's like this (isn't it?) … ; It's (not) such that …
Asserting a point[109]

Responsive, conclusive functions

d) *aisi baat nahi hai*
It's not like that. (lit. like-this thing not is)
Expressing disagreement / correcting a point

e) *aisa nahi (hai)*
It isn't like that! / Not like that!
Expressing disagreement / correcting a point

f) *aisa thodi (na) hota hai; aisa thodi na (hota) hai*
It's hardly like that! / It isn't like that at all!
Expressing disagreement / correcting a point

g) *aisa kuch nahi tha/hua*
It wasn't anything like that! / Nothing like that happened![110]
Expressing disagreement / correcting a point

h) *aisa (hota) hai*
That's what happens. / It's like that. / That's how it is.
Concluding a point or explanation

Fig 170: Prolific discourse-structuring expressions using *aisa* identified in the Pithoragarh data

The discourse-structuring expressions in Fig 170 are built upon the combination of similative demonstrative of quality *aisa* along with the copula. Different tenses and aspects of copula use are employed for different discourse-structuring functions. Optional words in brackets include the invariable question tag *na*, the copula auxiliary and the present tense conjugation of the copula meaning 'happen, occur'. The fixed expressions listed include a range of discourse-structuring roles from initiating a turn, specifically a longer turn where the expression *aisa hai na* 'It's like this' precedes an explanation or assertion, to expressions for agreeing and disagreeing and for concluding a topic. Notable is the expression in (d), *aisi baat*

---

[109] This can also be used as an interrogative to check understanding: *aisa hai?* 'Is that it? (= 'Is it like that'?)

[110] This phrase is also commonly used in the present tense but was not attested in the data set.

nahi hai 'It's not like that!', which is used extensively in colloquial conversation, especially to placate an upset party. It is only one of two expressions which employ another lexeme, and the only one in which *aisa* functions as an adjectival modifier of a noun (here *baat*) and not a pronominal sentence subject. The highly polysemous *baat* functioning here as a vague noun, reminiscent of Andersen (2010) where vague nouns in Norwegian including *ting* 'thing(s)' are used to structure discourse in similar ways, as in Fig 171:

> **tingen er** *jeg vil ikke ha hustelefon*
> '**the thing is** I don't want a house phone'
> **Fig 171: Copied from Andersen (2010:38, Fig 10)**

By contrast, the expression in (f) has a sense of indignance about it and employs an emphatic form of the quantifier *thoda* 'a little, a bit' with the negative marker: *thodi na*. This appears in other contexts as an adverb meaning 'hardly' and is often pronounced in elided form as *thodna*.

8.7 Part VI Functions of *type* in web data

8.7.1 Similarities with Pithoragarh data

The web data reveals a large number of *type* functions that are common to the spoken data, including categorization, phoric modification, similative comparison, hedging and quotation marking. These will be briefly illustrated with examples from the Hindi Web corpus.

Categorization

In a blog piece by one of Bollywood's finest female actors (Fig 172), Deepika Padukone is giving her views on changing Indian society and particularly how the definition of love is changing.

वे **पुराने टाइप के रोमांस** को ज्यादा प्राथमिकता नहीं देते।

*Ve **puraane type ke romance** ko jyaada praathmikta nahi dete*

They don't value the **old kind of romance** (lit. give priority to)

Fig 172: The thoughts of Deepika Padukone printed in the article
https://navbharattimes.indiatimes.com/-/articleshow/11876386.cms

This is a clear example of the categorizer usage of *type* where the adjective 'old' relates to the subcategory or kind of romance which is apparently not valued, and not to the category 'romance' in general. Other lexical items which modify the subcategory in the web data include quantifiers and cardinal numbers as well as determiners such as हर टाइप *har type* 'every type', as in हर टाइप के बैग *har type ke bag* 'every type of bag'.

Phoric modification

There are proportionally many fewer cases of determiners with *type*, but they are attested. They tend towards negative connotations, such as illustrated in Fig 173, a facet observed in the Pithoragarh data.

is type ke niraadhar aur kaalpanik wastuon ko kisi bhi jimmedaar media me jagah kaise di ja sakti hai?

How can any responsible media give space to **this kind of unfounded and imaginary subject-matter**?

Fig 173: From janokti.com

Fig 173 also illustrates how the phoric modifier takes on an instance reading: even without the full context of the comment it is clear that the writer is commenting on some specific subject-matter that s/he has read. Expressing his/her disapproval through the use of categorization, condemning the text at fault by placing it in a (sub)category of 'unfounded and imaginary subject-matter', lends more weight to the criticism.

Most of the similative uses of *type* are reminiscent of the ad hoc categorizer construction and could equally be construed as ad hoc categories. Like the ad hoc categorizer uses they typically consist of a named prototype or example followed by *type* and end with a very general category. *Type* is sometimes followed by the genitive, but is often omitted, as in Fig 174 in which famous cricketer Virender Sehwag is the named prototype or exemplar of an ad hoc subcategory of cricketers who have similar skills or features to Sehwag. However, the simplest interpretation is not the formation of an ad hoc category (based on similarity), but a similative comparison. In actuality, the difference between the two is minimal.

यहाँ **वीरेंदर सहवाग टाइप खिलाडी** पसंद किये जाते हैं
*yahan **virender sehwag type khilaadi** pasand kiye jaate hain*
Here **players like Virender Sehwag** are appreciated
Fig 174: From adityasi.blogspot.com

Although the majority of the examples fit the description above with a prototype to the left of the type noun and a general category to the right, the sentence in Fig 175 shows how the left and right nouns are reversible, using this strategy deliberately and ironically as if to suggest that it is in doubt which of the two is higher in the hierarchy. Here in place of a type noun, the phrase can also be translated using the English suffix *-like*: 'a *baabaji*-like *saadhu*, or should we say [a] *saadhu*-like *baabaji*'.

गाड़ी जब चली स्टेशन से तो **एक बाबाजी टाइप साधु** या कहें कि **साधु टाइप बाबाजी** ट्रेन पर चढ़ गये।
*gaadi jab chali station se to **ek baabaji type saadhu** ya kahen ki **saadhu type baabaji** train par chadh gaye.*
When the vehicle left the station then **a *baabaji* kind of *saadhu***, or should we say [a] ***saadhu* kind of *baabaji***, got on the train.
Fig 175: http://fursatiya.blogspot.com/2009/05/blog-post_13.html

The difference between the two nouns *Baabaji* and *saadhu* 'holy man' in India is one of status. Baabaji is an honorific given to certain holy men who are (more) revered. Those honoured as a *Baabaji* may dress in white, whereas *saadhus* all wear orange (see photo in Fig 176).

179

Fig 176: Saadhu Pushkar from Rajasthan in India

Similar to the spoken data, the web data showed a lack of manner usage. There were no cases of adverbial assimilation of *type*, nor of its use as a manner similative (either demonstrative or non-demonstrative).

Hedging

There is a high proportion of *type* tokens used for hedging, as in the spoken data, although only one of the hedging tokens could be identified as a hedge for mitigation, as in Fig 177, where *type* hedges the negative adjective 'worthless', perhaps to reduce the level of criticism of the Hindi character under discussion (छ).

यह छ है भी थोड़ा छिछोर टाइप का ,हर जगह जा घुसता है और क्ष को निष्कासित कर देता है।
ye छ *hai bhi thoda chhichhor type ka ,har jagah ja ghusta hai aur* क्ष *ko nishkaasit kar deta hai*
This छ is also a bit kind of worthless, it gets in everywhere and expels क्ष
Fig 177: From banarahebanaras.com; punctuation in original

In some cases, *type* may fulfil the dual function as a flag for code-switched items, although these number only 15 of 51, so it by no means explains the prolific usage of *type* with adjectival description.

Use with bracketed content

There were four instances in which *type* appeared in a short phrase inside brackets (), in each case providing additional information on the previous text, including a translation into Hindi such as in Fig 178 where the Hindi specifying what the speaker meant by *red spots* is given:

चकत्ता *chaktta* 'round blotch or mark (on the skin)' (OHED).

*earlier i was facing the skin problem after that this allergic asthama develops to me,In summers red-red spots(**chakkte type**) will start to come on the skin*
Fig 178: From aayushved.blogspot.com; text printed as written

Other examples include monolingual Hindi text (in Devanagari script) and mixed Hindi-English text, showing that this usage of *type* is universal regardless of language and script. The physicality of the written text allows us to perceive the typed brackets, yet there were similar uses of *type* from within the Pithoragarh spoken data, such as in Fig 179 where the phrase *daane-2 type ke* 'kind of granules' appears as additional information expounding the description of non-slippery tiles given immediately prior *dardare waale* 'coarse gravelly ones'. The brackets have been added to the utterance in Fig 179 to indicate a parallel between the usage in spoken conversation and in the written forum.

*dardare waale jo aate hain na (**daane-daane type ke**) vo itne khatarnaak nahi hain*
The coarse gravelly ones you can get (**kind of granules**), they aren't so dangerous
Fig 179: Anita (F32), convo_b

In both cases, the additional information is approximate, as indicated by the use of *type* functioning here as a hedge. And yet there is an element of semantic persistence as the presence of the type noun carries with it an element of categorizing or defining, albeit approximatively. This is reminiscent of the work of Vassiliadou et al. (forthcoming) who argue for a cline between categorizer and altogether vague uses of the binominal construction with French type nouns.[111]

Quotative

Similar to the Pithoragarh data, there were no instances of *type* marking actual reported speech. Rather, there were many and varied examples of quotative usage involve quasi-quotes such as the following: a well-known line of poetry; inner thought (one's own); hypothetical speech or song (someone else's); and a list of hypothetical story titles.

अगर दूर-दर [sic] तक सबकी गयी होती है, तो हम तसल्ली से बैठ जाते हैं **कि** एक हम ही नहीं अंधेरे में गालिब, जमाने भर में अंधेरा है, टाइप।
*agar dur-dar [sic] tak sabki gayi hoti hai, to ham tasalli se baithh jaate hain **ki** "ek ham hi nahi andhere me gaalib, jamaane bhar me andhera hai", type*
If we look further/deeper, then we'll sit comfortably [in the knowledge] **that** "**It isn't only us who are overcome in darkness, the whole generation is in darkness**", **like**.
Fig 180: From puranikalok.blogspot.com

---

[111] Although French *type* is less developed as an approximative than other type nouns.

An example of *type* marking one's own inner thought is given in Fig 180. The quotation is introduced by complementizer *ki* and the single-word quotative marker *type* follows the (hypothetical) quoted text. This usage of *type* is similar to that of Brazilian Portuguese *tipo* and due to the elision of the genitive represents a more advanced stage of grammaticalization than the quotative examples in the Pithoragarh data. In Fig 181 too, the quasi-quoted text, which is in fact a list or summary of occurrences, is not introduced by any quotative element, but marked with *type* following the text.

अब इस तुरंता माध्यम के चलते फ़ोटो की गति **खेंचा/देखा/भूल गये टाइप** हो गयी हैं।
*ab is turanta maadyam ke chalte photo ki gati **khencha/dekha/bhul gaye type** ho gaye hain*
Now with the speed and immediacy of taking photos while walking along, it's become **kind of, "Took photo, looked at it, forgot it"**
Fig 181: From hindini.com

Voghera describes the use of Italian *tipo* to mark such quasi-quotes as "labelling" (2013:297) rather than quotative marking, since *tipo* mostly highlights 'maxims, formulaic expressions, or pieces of common knowledge' rather than reported speech. The construction in which this usage appears in Hindi "REP + TYPE + N" is also similar to the Italian construction "N + TIPO + REP" which in Italian treats the quote as an adjective.

8.7.2   Differences from Pithoragarh data

No phoric similative or other extended uses of type demonstrative construction

Despite the tokens of phoric modification in the Hindi Web data, all cases of the demonstrative combining with *type* involved the proximal demonstrative *is* 'this' in the binominal construction "DEM + TYPE + GEN + N". There were no extended uses of *type* in combination with the demonstrative, notably not a single phoric similative. There were also no cases of demonstrative phrases in clause final or post-clausal position functioning as a general extender or filler or in any other discourse function. As a whole, the percentage of determiner phrases with *type* was small compared to the spoken data.

Person descriptor

One of the most striking aspects of the Hindi Web corpus is the high proportion of tokens in which a *type* construction is used to describe a person through comparison with a named example or application of an innovative adjectival descriptor. Some of these are hedges, but the majority are not. In fact, more than a third (59/148) of all uses of *type* appear in the

context of describing people, including eight tokens with a named person in the N1 position (before *type*), such as in Fig 182.

> ऐसे मे **हाशिए पे पड़े दिग्गी टाइप के नेता** भी बुकि बन कुछ रोजगार तो हासिल कर ही सकते है [*sic*]
> *aise me **haashiye pe pade Diggi type ke neta** bhi bookie ban kuch rojgaar to haasil kar hi sakte hai [sic]*
> In this way, **down on their luck politicians like Diggi** can become bookies and provide income for the unemployed.
> **Fig 182: From kalyugeenarad.blogspot.in**

In Fig 182 the *type* phrase makes reference to politicians showing similar characteristics to, and therefore in the same ad hoc subcategory of politicians as, the named exemplar Diggi. The role of *type* could either be understood as an ad hoc categorizer or a similative.

The following example is an ironic usage of the same *type* construction, which clearly evokes an ad hoc category (Fig 183):

> *aisa lagta hai ki **aasmaan se utare devdut type ke log** hain ye*
> Just as nothing negative is published / circulated against Mukesh Ambani, similarly not against Sonia-Rahul, it seems that they're **angels-descended-from-the-sky kind of people**
> **Fig 183: From blog.sureshchiplunkar.com**

It therefore seems that in the web data, *type* is used as a marker of personal description. It is not unexpected for a *type* cognate to develop a function in the area of personal description. There are two cross-linguistic observations that can be made in this regard. The first is that grammaticalization often occurs concomitant with a process of subjectification wherein speakers err towards increasingly evaluative use of language. Of course, evaluative statements can be made on a whole variety of topics, but subjective statements about people seem to be common contexts for type nouns to appear in. The Pithoragarh data shows this, with approximately a third of tokens used to comment on people's physical appearance, character or behaviour, often employing *type* to show similarity or hedged to mitigate possible bad feeling. There is only one example that may constitute an incipient usage of this 'person with certain characteristics' meaning, where participant Ganesh describes another person as an *upset type*. However, the meaning is ambiguous as a hedging interpretation 'kind of upset' is also possible. More strikingly, both Polish and Russian demonstrate a shift in the meaning of *type* to refer to a type of character, using metaphorical reference to movie or literary characters, e.g. *Polish typ allenowski* 'Woody Allen type'. *Type* also appears in a binominal construction, although this is more common for Russian, e.g. *vostocnyj tip lica* 'oriental type of face' (Kisiel and Kolyaseva, forthcoming). These structures

and usages in Slavic languages are a direct parallel to the many examples in the Hindi data which use metaphorical reference to a supposed 'type' or via a named person. Fig 174 is an example of such proper noun usage where famous Indian cricketer Virender Sehwag is named as a desirable 'type of cricketer'.

Secondly, there is a pattern of semantic shift attested in Italian, Polish and Russian, in which the *type* cognate shifts to a non-categorizer meaning of 'someone' (Voghera, 2013), or more specifically 'a guy', usually with negative connotations (Kisiel and Kolyaseva, forthcoming). Modern English also shows two even more restricted usages, both informal: 'A person of a specified character or nature' where type must occur with an adjective, e.g. 'two sporty types' (OED1.1); and 'The sort of person one likes or finds attractive' (OED1.2). It is possible that Hindi *type* is developing in this direction.

Syntactic differences

Type appearing alone without the genitive

From a syntactic point of view, the data sets show different preferences for which constructions *type* appears in, and particularly, whether the genitive is present or absent. In the spoken data from Pithoragarh we observe that the overwhelming majority of *type* tokens occur with the genitive attached, reminiscent of the binominal construction "[ADJ/DEM] + TYPE + GEN + N", whereas in the web data, the majority of constructions occur without the genitive, including 22 examples of the structure "N + TYPE + N". Since elision of the genitive (or other adposition) from type noun constructions is a feature of grammaticalization and well attested in languages (see Chapter 7), the higher percentage of *type* tokens without a genitive may be evidence of more advanced grammaticalization.

Type modifying phrases and clauses

There is also a token in which *type* fulfils a hedging function and could be translated either using the vague clause-final tag 'type of thing' or the hedge 'kind of' (Fig 184):

> किसी **यूनिवर्सल ट्रुथ** की तरह, जो बिना सिद्ध किये ही जैसा है वैसा ही रहेगा, **टाइप**.
> *Kisi **universal truth** ki tarah, jo bina siddh kiye hi jaisa hai vaisa hi rahega, type.*
> Like some **universal truth** which without being proven remains just as it is, **type of thing**.
> **Fig 184: From http://prashant7aug.blogspot.com/2012/03/blog-post_25.html**

Its placing at the end of the sentence is unusual and reminiscent of the most highly grammaticalized Italian uses of *tipo*, French *genre* and English *kind of*.

8.7.3 Conclusion

Whilst there are many similarities, the differences between the uses of *type* in the spoken Hindi data from the remote Lower Himalayan town of Pithoragarh and those identified in the informal written web corpus Hindi Web are prominent and warrant an explanation. This will be pursued in Chapter 10: Discussion.

8.8   Conclusion to synchronic functions of *type*

In this chapter, the synchronic semantic-pragmatic functions of loanword *type* in spoken Hindi have been described and analysed in the context of how other (pre-existing) Hindi lexemes fulfil the same functions. Considering the language-internal context has allowed us to observe how *type* seems to have become integrated into the lexicon by imitating other Hindi lexemes in various ways. For example, even as a categorizer noun, and in contrast to its function in English, *type* does not exhibit all the properties of a fully lexical noun, notably lacking the propensity to be pluralized. Like *aisa* and related lexemes from the similative demonstrative paradigm, *type* is also seen to have developed functions as a phoric similative, a hedge and a quotative and fulfils discourse functions as a general extender and filler. The phoric similative function is of particular importance as this innovation is largely unattested by type nouns in other languages, with Norwegian the only exception.

Extensive non-nominal uses of *type* appear to be synchronic evidence of a process of grammaticalization which has been underway within the past 100 years, and likely within a much shorter time span.

9   Grammaticalization of *type* following the model of *aisa*

9.1   Introduction

In the previous chapters I have used a comparison of research into *type* cognates cross-linguistically with the *type* data attested in Pithoragarh and supported by data from the Hindi Web corpus to argue that the various semantic-pragmatic functions of *type* and the syntactic contexts in which it appears seem consistent with the results of a process of grammaticalization. Chapter 8 clearly demonstrated that in each semantic-pragmatic context in which *type* appears in the Pithoragarh data, native equivalents *aisa* and its related lexemes (particularly *jaisa*), which pre-date *type*, are also attested. This chapter presents a reasoned response to the question of how *type* came to develop the innovative semantic-pragmatic uses attested in the corpus. Using a combined theoretical framework of language-internal grammaticalization in conjunction with De Smet et al.'s (2018) perspective on competition theory, notably the principle of attraction, a hypothesis for how *type* may have grammaticalized using native Hindi lexeme *aisa* as a model is proffered as the most reasonable explanation of the data.

Since this argumentation depends on the status of *aisa* as a grammaticalization model, a lexeme which has already been grammaticalized along a certain path and whose process can therefore be replicated, it is important to argue explicitly here what has been hitherto implied, that the Hindi similative demonstrative *aisa* and its relative lexemes (*aise, jaisa, jaise, vaisa* and *vaise*) show all the signs of having grammaticalized. This will be argued first before a full discussion of the process likely followed by *type*.

9.2   Grammaticalized *aisa* and *jaisa*

Very little has been written about similative demonstrative *aisa* and its morphologically related lexemes. Montaut (1995) introduces *jaisa* as a similative, yet focuses predominantly on the extended uses of *sa*, a prolific hedge derived from *jaisa* (see Chapter 8.5.1). The article does not comment on the etymology of *aisa* or the chronology of the grammaticalization of *aisa* vis-à-vis *jaisa*. However, based on Diessel's (1999) evidence-based claim that exophoric deixis is primary and that endophoric uses developed successively, it is reasonable to assume that *aisa* (or rather *aise*; see Chapter 8.3.3) preceded *jaisa*. Furthermore, it is logical that monomorphemic *aisa* would have existed prior to its morphologically more complex corelative *jaisa (j- + aisa)*. Indeed, following the line of argumentation pursued in this thesis which depends strongly on cross-linguistic semantic trends, phoric modifier uses of *aisa* in partnership with *jaisa* which have a strongly categorizing role (such as in Fig 185) are

expected to precede the development of independent similative uses of *jaisa* and the phoric similative function of *aisa*. A detailed diachronic study is needed in order to test this hypothesis.

> ***aisi*** *dukaan khol lo na phir! hamaari **jaisi***
> Open **that kind of** shop then! [**The kind**] **like** ours
> **Fig 185: Ganesh (M55), convo_t**

The pragmatic functions fulfilled by highly prolific *aisa* include some but not all of those detailed in König (2017) for similative demonstratives; functions such as quotative and booster were attested in the data, but not affirmative or additive. Examples of quotative usage in German and Finnish are reproduced in Fig 186a and 186b respectively. König (2017) points out that the proximal similative demonstrative is most commonly used for cataphora, which can then develop into the quotative usage, as illustrated by the Finnish example.

> a. *Und ich **so**, "…". Und er dann **so**, "…".*
>   'And I'm **like**… And he's **like**…'
> b. *Hän sanoi sen **näin**: "…"*
>   'He said it **like this**: "…".'
> **Fig 186: Reproduced from König (2017:160, examples 31c and 33)**

Also significantly, the high numbers and very frequent usage of fixed expressions with *aisa* in a discourse-structuring role mirror very closely the uses of German *so* (König, 2020), as described in Chapter 7.3. Although König's analyses are based predominantly on European languages, mention is also made of Japanese. König (2020) considers the question of whether the discourse-structuring functions of similative demonstratives should be regarded as the result of grammaticalization and concludes that the term is appropriate since the innovations demonstrate some of the key parameters of grammaticalization, including semantic bleaching, extension to new contexts, decategorialization and persistence, and that 'a concept of grammar that excludes dialogue-structuring devices would be too restrictive' (2020:31). This concurs with Heine's (2018) defence of the term for the development of discourse pragmatic markers.

The functions of *aisa* and *jaisa* display evidence of semantic extension to new contexts of use and change of syntactic status (from adjective to preposition to adverb), both parameters associated with the process of grammaticalization. For example, semantic extension seems to have led to a further development from similative to hedge, through semantically ambiguous sentences such as in Fig 187, where *halwa jaisa* could be equally interpreted as '(something) like *halwa*' or 'a sort of *halwa*'. This constitutes a bridging context.

> *inhone halwa **jaisa** bana rakha hai pura ghar*
> They've made the whole house **like** *halwa* (i.e. messy)[112]
> **Fig 187: Anita (F32), convo_b**

A hypothesis for the process of grammaticalization followed by *jaisa* is that having become established as a hedge, *jaisa* then starts colligating with adjectives, thus losing its adpositional status and becoming an adverb.

The pragmatic functions of *aisa* and its related lexemes and the syntactic realization of those functions are aligned with both the cross-linguistic tendencies highlighted by König (2020) and those common to the grammaticalization of similative demonstratives in other languages as identified in Chapter 7.3.

Therefore, on the basis of comparability with the grammaticalization of other similative demonstratives cross-linguistically, clues such as the layering of multiple semantic and syntactic uses within the synchronic data, and the persistence of either the deictic or the similarity meaning on the part of *aisa*, it seems reasonable to conclude that *aisa*, *jaisa* and their counterparts have indeed grammaticalized.

Table 3 illustrates the semantic-pragmatic functions of *aisa* and its (derivationally) related lexemes in the context of the functions of the other similative demonstratives studied in Chapter 7.3. The rows in purple show the functions fulfilled by each lexeme separately. Considered individually, this may present a confusing or disjointed picture, and yet when the lexemes are viewed together, as in the pink row, it becomes clear that the expected innovative functions of a similative demonstrative (based on the cross-linguistic study) are shared between *aisa* and its (derivationally) related lexemes *jais*a, *sa, vaisa*, *aise*, *jaise* and *vaise*, which have largely specialized in terms of their semantic-pragmatic roles. In the context of the discoveries made about similative demonstratives thus far, this is unprecedented and groundbreaking. Bar Chart 2 provides a visual depiction of how the semantic-pragmatic functions are divided between the lexemes, based on a subset of the corpus which incorporates speakers of both genders and multiple age groups. The six recordings included are: convo_a, convo_b, convo_d, convo_e, convo_g, and convo_n. (See Appendix C for details.)

What is of paramount importance for this thesis is that all of the innovative semantic-pragmatic functions of *type* are mirrored in well-established functions also fulfilled by *aisa, jaisa* and related lexemes. The much longer history that these lexemes have enjoyed in the

---

[112] *Halwa* is a common dessert made from shredded carrot boiled for a long time with sugar. Although it tastes delicious, it looks messy, hence the metaphorical extension.

language (having been adopted from Sanskrit), coupled with their much higher frequency of occurrence in the data, lead to the conclusion that the innovative uses of *aisa* and *jaisa* described in this thesis are better established than those of *type*. This seems to have allowed the older lexemes to become a grammaticalization model for *type* through the process of attraction, a line of argument which will be pursued in the following section.

Language	Word	Phoric modifier	Phoric similative (quality)	Phoric similative (manner)	General extender	Intensifier	Similative (quality)	Similative (manner)	Exemplifier	Hedge	Quotative	Filler	Focus marker	Other discourse function
English	*such*	(✓)			✓	✓	-	-	✓					✓
English	*so*			(✓)	✓	✓	-	-		✓				✓
German	*so*	✓	✓	✓	✓	✓	-	-		✓	✓		✓	✓
Danish	*sådan*	✓	✓	✓	✓	✓	-	-	✓	✓	✓	✓	✓	✓
Swedish	*sån*	✓				✓	-	-		✓	✓		✓	✓
Norwegian	*sånn*	✓	✓	✓		✓	-	-		✓	✓	✓	✓	✓
Brazilian Port.	*assim*	✓	✓	✓			-	-	✓	✓	✓	✓	✓	✓
Hebrew	*kaze*	✓		✓	✓		-	-		✓	✓		✓	
Estonian	*nii*			✓			-	-			✓			✓
Hindi (separate)	*aisa (vaisa)*	✓	✓	(✓)		(✓)					✓	(✓)	(✓)	✓
Hindi (separate)	*aise*			✓	(✓)	(✓)					✓	(✓)	(✓)	✓
Hindi (separate)	*jaisa*	(✓)	✓				✓			✓				
Hindi (separate)	*sa*				✓					✓				
Hindi (separate)	*jaise*						✓	✓						✓
Hindi (separate)	*vaise*			✓										✓
Hindi	*aisa etc.*	✓	✓	✓	(✓)	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 3: Semantic-pragmatic functions of *aisa* and related lexemes from a cross-linguistic perspective
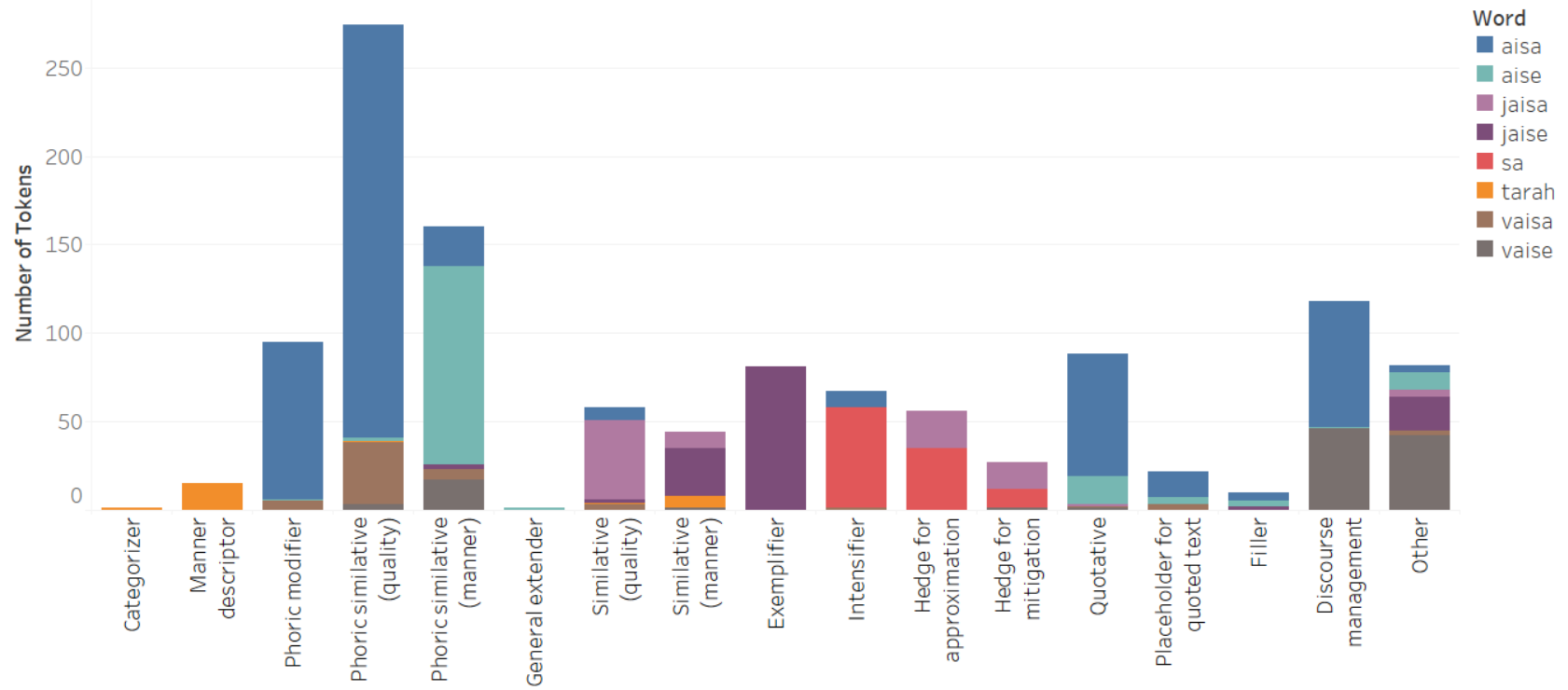
Bar Chart 2: Semantic-pragmatic functions of *aisa* and its related lexemes

9.3   Motivated argument for the grammaticalization of *type*

Having established the strong likelihood that *aisa* and related lexemes have undergone a process of grammaticalization (which corresponds with the development of similative demonstratives cross-linguistically), our focus turns to *type* and its journey of semantic innovation and grammaticalization. Table 4 illustrates the way in which Hindi *type* fulfils many of the semantic-pragmatic functions of *type* cognates cross-linguistically. The only notable exception is the exemplifier function, which is fulfilled almost exclusively by the adverb *jaise* in Hindi. The fact that *type* has not (yet) developed into a focus marker is not surprising, since the focus marker function is one of the most grammaticalized. A number of the functions were identified as incipient. For example, there was only one token of a *type* construction being used as a filler. The lack of tokens of *type* fulfilling functions towards the right hand side of the table is not an indication of it not having grammaticalized, but simply not having grammaticalized fully. However, the fact that spoken *type* data seems to mirror the development of *type* cognates in so many other languages, particularly languages with which Hindi has (had) little to no contact, is a strong indication that the cognitive processes affecting *type* globally are universal.

We have also noted the many similarities between the development of innovative functions of *type* cognates and those of similative demonstratives, both cross-linguistically and within the Hindi data. Bar Chart 3 makes this even more explicit. The bar chart shows the total number of tokens from the same six recordings as in Bar Chart 2. These six recordings account for approximately half of the tokens of the target lexemes in the corpus, including 44 out of the 67 *type* instances (66%) and 533 out of 700 *aisa* instances (76%). We can observe the way that *type* appears throughout the chart, except at the far right hand side, registering uses in the same functional areas as *aisa* and *jaisa* in particular.

As Heine and Kuteva (2005, 2010) and others have shown, the process of grammaticalization and the pathways along which linguistic items develop cross-linguistically apply equally to the language contact environment as to the language-internal environment. The environment in which we find *type* as an established borrowing in Hindi could be considered a hybrid between these two environments, or perhaps a novel third environment, incorporating language contact and yet being a language-internal process. We noted in Chapter 4.6 that within the lexicon the principle of attraction is at work, causing one lexeme to adopt characteristics of another via analogical comparison and replication. What follows is a theorized account of how *type* may have come to be attested in the constructions and with the pragmatic functions that are evident from the Pithoragarh data.

Language	Word	Categorizer	Phoric modifier	Phoric similative (quality)	Attributive modifier	Ad hoc categorizer	General extender	Similative (quality)	Exemplifier	Hedge	Quotative	Focus marker	Filler
English	*type*	✓	✓		✓	✓	(✓)			✗	✗	✗	
Norwegian	*type*	✓	✓	✓	✗	✓	✓		(✓)	(✓)	(✓)	(✓)	
Norwegian	*typ*	✗	✗		✗	✗	✗		✓	✓	✓	(✓)	
Swedish	*typ*	✓				(✓)			✓	✓	✓	(✓)	
French	*type*	✓	✓		✓		✗		✓	(✓)			
Italian	*tipo*	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	
European Portuguese	*tipo*	✓	✓		✓		✗	✓	✓	✓	✓	✓	✓
Argentinian Spanish	*tipo*	✓	✓			✓	✓	✓	✓	✓	✓	✓	✓
Russian	*tip*	✓				✓	✓	✓	✓	✓	✓		✓
Polish	*typ*	✓				✓		✓	✓	✓	✓	✗	
Hindi	*type*	✓	✓	✓	✗✓[113]	✗✓	(✓)	✓	✗	✓	✓	✗	(✓)

Table 4: Semantic-pragmatic functions of Hindi *type* from a cross-linguistic perspective

---

[113] Both here and in the adjacent cell ("Ad hoc categorizer"), the cross relates to the spoken corpus and the tick relates to the Hindi Web data.
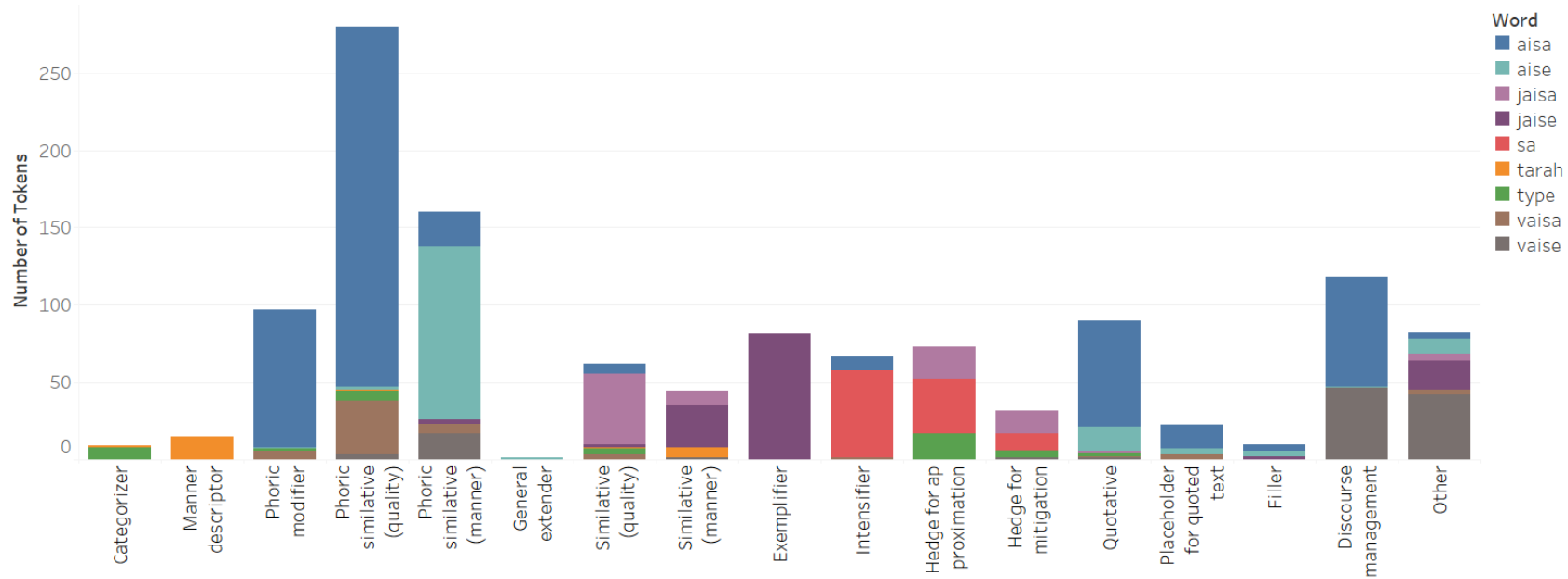
Bar Chart 3: Semantic-pragmatic functions of *aisa* and *type*

9.3.1   Summary of argument

In Italian, the lexeme *tipo* had existed with the closely related meanings of 'model', 'prototype', and 'exemplar' from the sixteenth century (Mihatsch, 2016; Voghera; 2013), but developed the classifier or categorizer meaning in the second half of the nineteenth century, as in English, whose first attested categorizer usage was in 1854 (OED). Given the language contact environment in India in the early twentieth century, specifically in Pithoragarh, where contact with English speakers was limited to three missionary families (who likely also spoke Kumaoni) and returning soldiers, it is not likely that loanword *type* would have become established in the lexicon quickly. It seems more likely that the rise in English-medium education at the end of the last century and particularly the beginning of this century would have been the catalyst for the propagation of *type*, as argued in Chapter 3, if not also for its initial borrowing. A hypothesis for the borrowing and grammaticalization of *type* is proposed in a step-by-step description in what follows.

1. *Type* is borrowed by (non-fluent) bilinguals who perceive it as an alternative categorizer noun to Sanskrit-origin *prakaar* and Arabic-origin *tarah* and start to use it in a similar context to classify or subcategorize other nouns. Being understood as an equivalent to *prakaar* and *tarah*, it is assigned the same syntactic structure and notably (almost) never used as a plural. (Either *type* never displayed fully nominal properties, or the plural dropped out of usage due to analogy with *prakaar and tarah*.)

2. Speakers replace *prakaar* with *type* in informal contexts, relegating *prakaar* to formal contexts congruent with its etymology as a Sanskrit word. *Tarah* and *type* occupy a similar semantic space as categorizer nouns, but the older loan *tarah* having developed a manner meaning, *type* and *tarah* fall into a pattern of semantic complementary distribution, one of the outcomes of competition theory (De Smet et al., 2018).[114]

3. Simultaneous to these developments, speakers motivated to use *type* look for opportunities to bring it into their everyday language and, noticing that *type* constructions with the demonstrative share some semantic-pragmatic contexts with similative demonstrative *aisa*, they seek to apply *type* to an additional context in which *aisa* is found (as per Matras, 2010). In doing so, the use of *type* is extended to

---

[114] Although 'complementary distribution' usually describes the alternation of different phonemes in different phonological contexts, the term is employed here analogically for lexemes which function in different semantic-pragmatic contexts.

the new context of the phoric modifier, rendering its syntactic status ambiguous. This extension follows the principles of grammaticalization including unidirectionality.

4. The use of a phoric modifier with a dropped category noun facilitates the proliferation of the construction 'DEM + TYPE + GEN' in other syntactic-pragmatic contexts, including as a phoric similative with the copula, and leads to its reanalysis as a similative demonstrative which can be used adjectivally and pronominally.

5. The resulting ambiguity is the beginning of an incipient category formation as reanalysed *type* constructions with both proximal and distal demonstratives progress towards similative demonstrative status along the model provided by *aisa*.

6. The similative demonstrative construction *is type ka* 'this/that type of' is now an equivalent of *aisa* in more than one functional and syntactic context, and thus in competition with it for use by speakers in those contexts. The principle of attraction, which operates on competing lexemes (De Smet et al., 2018), creates a cognitive demand for *(is) type (ka)* and *aisa* to become increasingly similar, and speakers, consciously or subconsciously, start to utilize *type* in more of the semantic-pragmatic contexts in which *aisa* is used.

7. Having assumed the syntactic identity of a complex similative demonstrative through attraction to *aisa*, *type* analogically patterns itself further on *aisa* by taking on a similative function fulfilled by the associated relative adjective *jaisa*. Note that since the semantic shift from categorizer to similative is cognitively motivated and extremely common cross-linguistically, this development would have been expected regardless of *type* becoming a similative demonstrative. However, in the Pithoragarh data demonstrative expressions are highly prevalent and so it seems likely that the phoric similative 'like this' meaning developed first, and the similative 'like' sense developed subsequently.

8. In addition to its development as a similative demonstrative (with adjectival status), *type* also develops as a similative *type ka* 'like', a postpositional role following nouns. Once established with the semantic-pragmatic function of a hedge (the relative timing of which is unknown), *type* then proceeds along the same pathway as *jaisa*, colligating with adjectives as well as nouns, with the same approximative meaning.

9. Via analogy with *aisa*, the *type* similative demonstrative complex *is/us type ka* also begins to function pronominally and its semantic-pragmatic meaning increases in both vagueness and in the possibility of negative connotations or euphemism.

10. This leads to uses as a quotative, a discourse marker and a general extender (via the adverbial manner demonstrative). These uses were infrequently attested in the data, suggesting they were incipient uses at the time of data collection. *Type* thus joins an illustrious collection of type nouns which have followed this path, including *type* cognates in Italian, Swedish and Russian (Voghera, 2013; Rosenkvist and Skärlund, 2013; Kolyaseva and Davidse, 2018).

9.4   Evaluation of alternative hypotheses

The process of grammaticalization presented above is motivated by the colloquial spoken data from Pithoragarh, supported by web data and, crucially, informed by cross-linguistic research. There are a few additional questions that arise on the basis of the theory proposed in the previous section, and these are addressed here. Furthermore, alternative hypotheses for how *type* may have grammaticalized, namely, language-internal grammaticalization with no analogical replication, are also shown to be less plausible.

9.4.1   Alternative hypothesis: Language-internal grammaticalization alone

Would type have developed all the same extended pragmatic functions and extended into all the same syntactic environments had *aisa* not existed in the language as a model? Of course, we can never know. However, we have recourse to cross-linguistic data such as those in Table 4 which suggest that the development of *type* into a quotative, discourse marker and general extender is part of the normal and expected language-internal grammaticalization process. However, certain unique features of the grammaticalization of *type* in Hindi are not reminiscent of the grammaticalization attested elsewhere. Firstly, the development of a *type* construction with the demonstrative into a similative demonstrative is unique globally, as far as research conducted to date has shown, with the exception of Norwegian *den type(n)* 'that type' which functions as a similative demonstrative in some, but not all, contexts. Secondly, the speed with which *type* seems to have developed so many non-categorizer functions and be employed in so many non-nominal contexts is surprising. The important question is whether the theory that *aisa* was a model for grammaticalization is more plausible than the alternative, a process of grammaticalization without a model. Regarding the speed of innovation, one could venture that the fashion of employing English words and the sudden explosion of English-medium education in Pithoragarh led to a rapid and prolific use of *type* that fuelled its equally rapid progression from categorizer to discourse marker. Indeed, the sociolinguistic prestige of the language must have played a role. Moreover, one could contend that the development of a similative demonstrative

construction is compatible with the categorizer-to-similative cognitive progression discussed earlier. Could it be, therefore, that the overlap with *aisa* is a coincidence? Or is it more likely that the presence of a polysemic and abundantly used partial synonym (*aisa*) has had an influence on the proliferation of *type* functions? Whilst the possibility of a coincidence cannot be completely refuted, the uniqueness of *type* taking on a similative demonstrative function combined with its speed of innovation indicates that the hypothesis of *aisa* as a grammaticalization model for the development of *type* seems more likely.

This being so, the role of the speaker and hearer in the process of change need to be reexamined. Chapter 5.3 described the creative role of both speaker and hearer which applies to the borrowing process and to the grammaticalization process, regardless of the trigger. Fig 188 outlines the agentive role played by speaker and hearer in the development of *type* in Pithoragarh Hindi. The borrowing process is summarized very simply in Steps (i) and (ii). The grammaticalization process is based on that presented in Fig 15, with the addition of Step (b) and modifications to Step (c), which are unique to the grammaticalization context presented here, the **native-lexeme-as-model** hypothesis.

<u>Borrowing process</u>

i.   initial borrowing as categorizer noun --> the "innovator" (a bilingual speaker)
ii. propagation throughout the community as categorizer --> all speakers (bi-/monolingual)

<u>Grammaticalization process</u>

(a) use of the lexeme in a semantically ambiguous context --> any speaker or innovator
(b) **identify analogy with native lexeme** --> any speaker/hearer
(c) semantic reinterpretation and syntactic reanalysis **based on polysemy of native lexeme** --> any speaker/hearer
(d) use of the new meaning in a new syntactic context --> any speaker
(e) propagation of the new form/meaning throughout the community --> all speakers

---

Fig 188

The hypothesis of grammatical replication involving *aisa* as a model for *type* depends crucially on Step (b), wherein the speaker or hearer (or both) identifies the lexeme *type* with the lexeme *aisa* in a particular context. As De Smet et al. claim, 'By virtue of their mutual similarity, functionally overlapping forms tend to be treated similarly [by speakers] and therefore may become more similar over time' (2018:227). In order for *type* to be identified with *aisa,* the context required is a *type* construction with the demonstrative, either proximal *is type ka* 'this type of' or distal *us type ka* 'that type of'. The syntactic bridge is the phoric

modifier construction with the dropped category noun which is reinterpreted as a predicative construction with the copula, leading to its function as a phoric similative.

Heine (2012:132) states that 'replica categories are generally less grammaticalized than the corresponding model categories'. If we understood English *type* to be the model, then we would not expect Hindi *type* to be as far along the grammaticalization process as English *type*, whereas on the contrary, we know that the Hindi 'replica' has developed many meanings that English *type* has not. This entails that English cannot be the model for grammatical replication and strengthens the argument proposed here that the loanword, having been borrowed as a categorizer noun, has developed via a process of language-internal grammaticalization in Hindi. The proposition that *aisa* is the grammaticalization model has already been mooted and several arguments in favour of this hypothesis have been presented. The account of events with the most explanatory power is undoubtedly the argument for language-internal process involving replica grammaticalization between *type* and *aisa.*

9.4.2 Alternative hypothesis: External influence via language contact

Another plausible explanation may be continued influence from English. One distinction between a loanword and a 'native' lexeme is that a loanword also retains the possibility of being borrowed again or hypercorrected by speakers with higher English proficiency or awareness. It is unlikely that *type* was initially borrowed with a categorizer meaning and later reborrowed with subsequent meanings. As Table 4 shows, binominal uses of *type* as categorizer and phoric modifier are the only functions that are shared by British English and Pithoragarh Hindi. The other attested usages of *type* in Hindi cannot have been borrowed, since they do not exist or are not well established in the source language.

Global influences in communication have led to more vague discourse and the growth of new quotatives, to mention two recent linguistic phenomena. These have been proposed as an explanation for youth language in other countries imitating the language of youth in America. However, as clarified in Chapter 3.3, the sociolinguistic environment of Pithoragarh at the time of data collection was such that few people actively engaged with western society through individual contact or TV and film. Kolyaseva (2018:86) responds to the claim of contact-induced influence of English *like* on Russian *tipa*, which has developed many of the same functions, concluding that due to a distinct lack of contact between English and Russian

during the relevant period, the parallel developments must be understood as 'representatives of the same crosslinguistic trend' and not a result of contact.

Sociolinguistic influences on grammaticalization are notoriously difficult to prove. The debate printed in the Journal of Sociolinguistics in 2014 tackling the issue of media influence on language change is evidence of this: scholars were fiercely divided with strong opinions on both sides (see Sayers, 2014, for the focus article and other articles in the same issue for the debate, plus a rejoinder from Sayers).

Considering the language contact argument, Mihatsch (forthcoming a) highlights various reasons why contact is less likely to be a causal factor, either because there was in fact no contact (Russian and Polish) or because the dates do not align, although of course a slight influence cannot be ruled out. Mihatsch (forthcoming a) also suggests that 'general international communicative tendencies' may have been a factor, although for this to be true, international communication of some sort must have been achieved, either personally or via the media, something which has not been prevalent in Pithoragarh, as discussed in Chapter 3.3.5.

9.5   Functional specialization between *type* and other lexemes

We have seen how the principle of attraction is at work in the Hindi lexicon. Other outcomes traditionally explained by the competition theory (or extended competition theory as per De Smet et al., 2018) are also in evidence. Within the extended uses of *aisa* and its many related lexemes (*aise, jaisa, jaise, vaisa* and *vaise*) there is a huge amount of scope for further study and even conclusions reached in the course of this study which for reasons of space cannot be expressed here. In this section, we will focus on the relationship between *type* and pre-existing Hindi lexemes. A further look at the distribution of lexemes between various semantic-pragmatic functions illustrates how their prolonged coexistence within the same semantic field seems to have resulted in two effects of competition, specifically substitution and differentiation.

9.5.1   Overview of lexeme functions

Bar Chart 4 illustrates again the number of tokens that occurred in the corpus for each pragmatic function attested by *type* and gives a detailed breakdown of how these were fulfilled linguistically using each of the target lexemes in the study. As in the previous chapter, pragmatic functions are arranged from left to right from most conventional to most innovative usage.
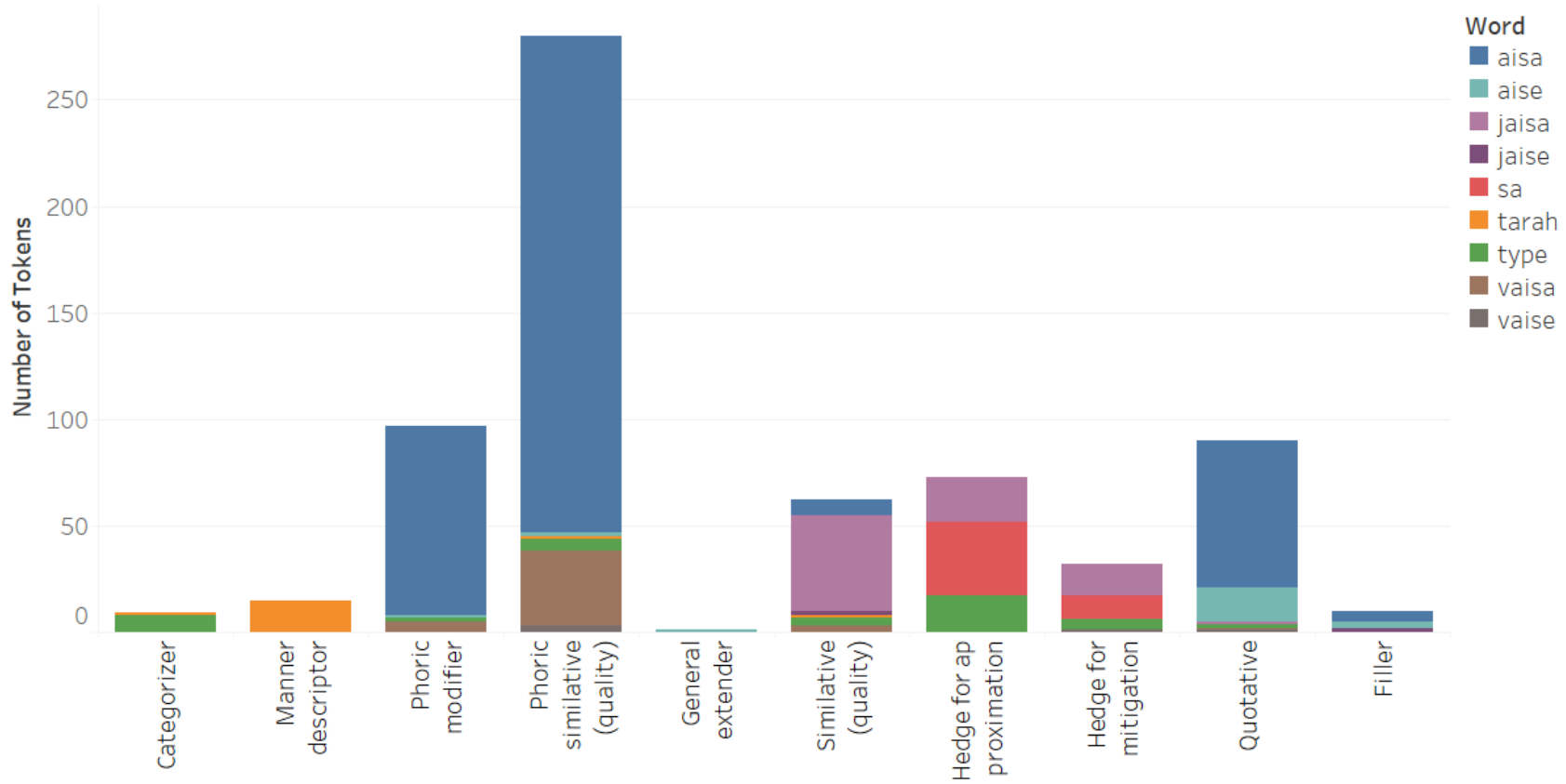
Bar Chart 4: No. of tokens for each function of *type*

At a summary level, the chart highlights several key differences in the usage of each lexeme. Key findings are as follows:

- *tarah* and *type* share the functions of categorization and manner description; *type* is more commonly used as a categorizer and *tarah* as a manner descriptor
- *prakaar* is barely used in this informal register
- *aisa* is prolific as a phoric modifier and phoric similative (of like quality), along with its contrastive corelative partner *vaisa*
- *jaisa* is by far the most frequently used similative (of like quality); note that *tarah*
- *jaisa, type* and *sa* all function as hedges
- *aisa* (and to a lesser extent *aise*) is used extensively as a quotative

Since the bar chart shows only the functions fulfilled by *type* it omits the similative of manner, exemplifier and discourse management functions, which is why *jaise* and *vaise* barely feature.

Two areas of functional differentiation between *type* and *jaisa* have been identified. Both lexemes function as similatives and hedges, and yet within these broad semantic uses more detailed analysis has revealed that *jaisa* is the preferred choice for metaphorical comparison, while the lexemes fulfil different functions as flags for code-switches. These will be expanded on below.

9.5.2 Metaphorical reference

Type nouns are polyfunctional, and in several examples *type* and its equivalents, particularly *jaisa*, function as markers of metaphor in addition to their other role in the utterance. There are 23 examples of metaphoric reference marked with *type* and its equivalent lexemes in the data, spoken by 12 out of the 47 participants. The table in Fig 189 shows that *jaisa* is the most common lexeme used for marking metaphor, accompanying proportionally more metaphoric references than *sa, tarah* or *type*. A total of 14 (12%) *jaisa* tokens are used to express metaphor, of which 13 left hand (LH) collocants are unique.

Metaphorical reference	Word				
	jaisa	*sa*	*tarah*	*type*	**Total**
Metaphor tokens	14	4	3	2	23
Total tokens	120	206	65	67	463
% Metaphor	**12**	**2**	**5**	**3**	**5**

Fig 189

The majority of metaphorical references appeared in the context of similatives and hedges. Bar Chart 5 illustrates the number of tokens of each of our target lexemes appearing in these contexts. Since metaphoric reference is a semantic extension of the use of a word, it is understandable that the most common pragmatic context is similative comparison (of like quality). However, metaphor is also used in veiled insults (alongside a hedge for mitigation) and expressive description (with a hedge for approximation).
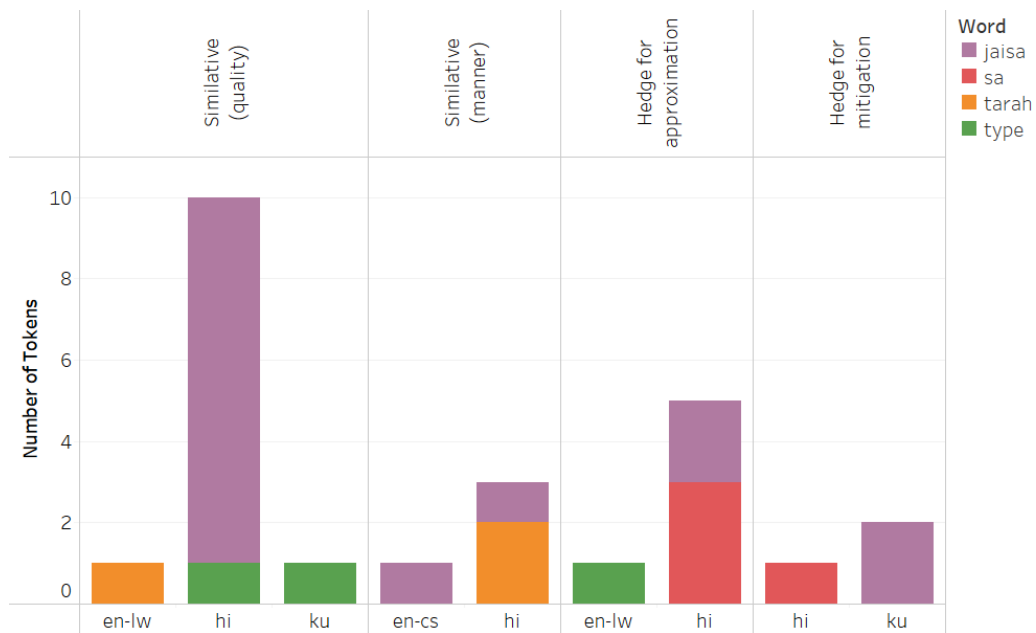
Bar Chart 5: No. of tokens of marking metaphor by lexeme and LH language

The bar chart also indicates the language of the word appearing immediately before the marker, i.e. the object being compared (for similatives) or hedged. Some of the metaphorical phrases are highly innovative and several use Kumaoni words. Bar Chart 5 illustrates how Kumaoni words are used with *jaisa* and *type*, but not either of *sa* or *tarah*. The single Kumaoni word used with *type* (*dokka*, a kind of basket; see Fig 126 in the previous chapter) is a common household object in all villages surrounding Pithoragarh and would be known by the majority of Pithoragarh residents.

As well as featuring highly as a similative, *jaisa* also appears twice in a hedge for mitigation, in the phrases in Fig 190 below. A *sintaalu* (Fig 192a) is a kind of bird found locally in Kumaon which has a long beak. Here it is used meronymically to refer to the boy who visited the house, also designated by *langur* 'langur', a kind of long-tailed monkey.

a. *vo jo langur nahi aata hai vo, **sintaalu jaisa***
   'that *langur* who comes [to the house], **like a *sintaalu***'
b. *photo me ajeeb hi aayi hai, **jhhultaryaani jaisi***
   She's come out weird in this photo, **kind of badly dressed**, hasn't she?

Fig 190

The second sentence contains a derogatory adjective also hedged by *jaisa* (*jaisi* with feminine marking). In Kumaoni the adjective *jhhultaryaani* (Fig 190b) typically describes objects that have been covered in smoke from an open wood fire (*jhhula*). The highly expressive metaphor when applied to a person is a comment on their appearance; they may be nice-looking, but are wearing torn clothes and have not washed their face, etc. These examples indicate that *jaisa* is the preferred collocant for Kumaoni words, particularly an idea further pursued in the following section.

9.5.3 *Type* as flag for English-origin words and *jaisa* as flag for Kumaoni-origin words

In Chapter 8.5.2, the use of *type* to flag less entrenched English-origin words (termed 'code-switches' here) was discussed. A comparison between collocants of *type* and *jaisa* when functioning as a hedge shows that *type* is more likely to be the hedge of choice following English code-switches, whereas established loanwords like nouns *operation* 'medical operation' or *form* '(paper) form' were equally likely to be hedged by *type* or *jaisa*. The data shows that of eight potential loanword collocants, three were followed by *type* and three by *jaisa*. Another key distinction between *type* and *jaisa* constructions is the language of the left collocant, specifically the likelihood of an English-origin word to form part of the construction. We see from the table in Fig 191 below that English code-switches are eight times more likely to be followed by *type* than *jaisa*. Therefore, the status of the English-origin words as 'not yet established' is a fairly accurate predictor of a *type* construction.

Language	*type* # words	*jaisa* # words
Hindi	20	31
English (all)	11	4
- code-switches	8	1
- loanwords	3	3
Kumaoni	1	11
Total	**32**	**46**

Fig 191

Just as *type* was seen to be the strongly preferred hedge for flagging English code-switches, we can see that the reverse is true for Kumaoni words. Bar Chart 6 shows that all Kumaoni words were followed by *jaisa*, although one word, *adiyaat* 'idiotic', was used with *type* and

jaisa by two different speakers respectively. A comparison of the lexeme choice of those speakers shows that this was consistent with their preference for one or the other lexeme overall. Anita, who used *adiyaat type ka*, showed a predilection towards *type* (11 tokens versus 15 for *jaisa*) whereas Vimla, who used *adiyaat jaisa*, showed a strong preference for *jaisa* in her overall use of the lexemes (6 for *type* and 25 for *jaisa*).
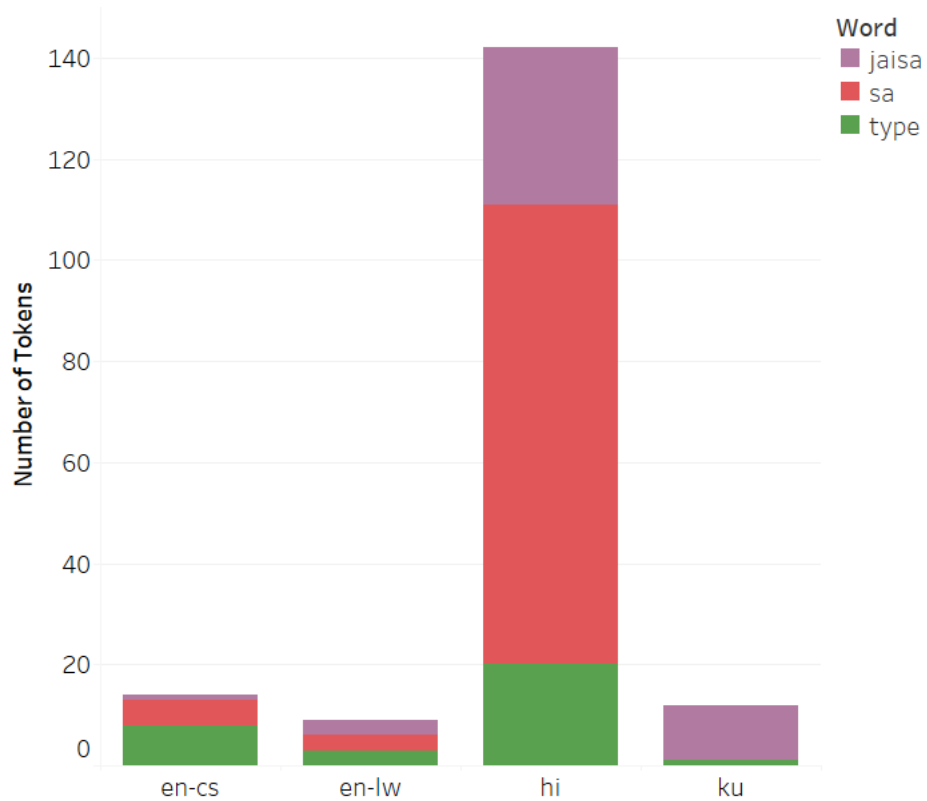
Bar Chart 6: Hedges by LH language by lexeme

We have established, both through the disclosure of a participant that English code-switches are likely to be hedged, and from the analysis in Fig 191 can conclude that these hedges are likely to take the form of a *type* construction. The previous chapter suggested two reasons why this might be so, from the speaker point of view and the hearer point of view. From the speaker point of view, a hedge allows the speaker to express uncertainty that they are using a less familiar word in the appropriate way. From the hearer point of view, the flagging of foreign words in conversation highlights to the hearer that the word may be less familiar to them, and primes them to expect a word of foreign origin.

By this subconscious process, selection of a less established English word automatically triggers an English-origin hedge, *type*, whereas the choice of a Kumaoni word triggers a Hindi-origin hedge, *jaisa*. Since established loanwords have become so integrated into the Hindi lexicon as to be perceived as Hindi (and no longer foreign), the priming effect would either

not be in effect, or would be significantly reduced, which explains why loanwords are equally likely to be followed by *type* or *jaisa*.

9.6   Conclusion

This chapter has been dedicated to an explanation of how the English loanword *type* may have grammaticalized in Hindi, drawing on several complimentary theories and inputs: grammaticalization, the principle of attraction (related to competition theory), universal cross-linguistic tendencies, and the role of the speaker and hearer in semantic change. As *type* is introduced into the lexicon as a partial synonym of *aisa*, it begins the process seemingly followed by all *type* cognates, that of progression towards similative, approximative (hedging) and extended discourse uses. Simultaneously, it also finds itself in competition with native lexemes. It was argued that the universal principle of attraction has explanatory power: *type* is attracted to partial synonym *aisa*, and replicates the grammaticalization process previously followed by *aisa*, which causes one departure from the expected pathway of development of type nouns (the phoric similative). It is the sociolinguistic motivation of speakers to use English words such as *type* that facilitates this process and likely increases the velocity of semantic innovation and grammaticalization.

10 Discussion

Having ascertained that *type* shows all the signs of having grammaticalized according to cross-linguistic expected norms, at the end of this thesis there is now an opportunity for meta-discursive commentary on the subject of cross-linguistic comparison, an integral part of the methodology employed throughout and central to the argument for the grammaticalization of Hindi *type*. In this chapter we also consider the importance of analysing *type* in the context of other lexemes fulfilling the same functions in the Hindi lexicon. Innovative functions of Hindi *type* not attested elsewhere are highlighted and finally the issue of spoken versus written data is considered.

10.1 Benefits and challenges of cross-linguistic comparison

In recent years the incredible similarities observed between the functions of grammaticalized type nouns from a range of languages has prompted an increase in focused comparison between them, notably by Mihatsch (2016; forthcoming a and b). Although causal links to global sociolinguistic trends and media influence have been proposed, they are partially explanatory at best and in some cases, for example, Russian, have little to no explanatory power (see Chapter 9.4.2). This strongly indicates that there are universal features of type nouns and demonstrates the huge benefit of conducting cross-linguistic comparison: there seems to be something about the nature of type nouns in general that leads to them developing a highly uniform set of functions. In focusing specifically on the lexeme *type*, this study highlights the veracity of this claim for *type* cognates in particular.

Despite the obvious benefits, there are also several challenges associated with cross-linguistic comparison: on the one hand, the focus on syntactic structures in many publications which makes cross-linguistic comparison for languages with inherently different syntax problematic; and on the other hand, the difficulties in defining discreet functions of type nouns without recourse to syntax.

It is a corollary that in order to conduct a meaningful comparison between linguistic elements in any two (or more) languages, the units of comparison must be of like kind. Each stage of grammaticalization is by definition characterized by both a novel pragmatic usage and a novel syntactic structure. Yet the data from the studies surveyed in this thesis show that there is rarely a one-to-one mapping between a semantic-pragmatic function of a word and its syntactic realization but a significant overlap of, in some cases, words or constructions with the same meaning, and in other cases, a single linguistic form with multiple meanings

(polysemy). For example, syntactic ambiguity in Swedish is such that a prepositional phrase with *typ* could mean 'type' or 'such as', and in turn the exemplifier meaning of *typ* 'such as' can be realized with either a prepositional or an adverbial construction (Rosenkvist and Skärlund, 2013). This dichotomy means that the identification of the semantic-pragmatic function of a word or construction must be considered independently from its syntactic realization and has precipitated a choice for scholars between focusing the unit of study on either the lexeme or construction on the one hand, or the semantic-pragmatic function on the other. Although semantic meaning is always considered, and in some detail, for the most part the full-length studies of type nouns available in the literature have tended to focus primarily on the linguistic form of a lexeme and its syntactic constructions. The principal problem with this approach is that different languages have inherently different syntax. As a very basic example, Hindi has postpositions whereas the majority of European languages have prepositions. Whilst this small issue can be overcome by using the descriptor 'adposition', there is also no guarantee that adpositions are functionally equivalent in all languages. Once syntax becomes the prime basis for comparison, we face thorny issues relating to the comparability of syntactic categories which need to be overcome.[115] The deliberate decision to designate the semantic-pragmatic function as the unit of comparison in this thesis is one such attempt. This decision also emanates from a sociolinguistic focus on the speaker and their message as of prime importance in analysing communication. However, the focus on semantic-pragmatic functions and the effort to disentangle the function from the form has led to other challenges.

Since many of the meanings and uses of *type* (and type nouns in general) have been defined in the literature on the basis of specific syntactic requirements, it has been necessary to create new descriptive categories. At times this has come into conflict with an endeavour not to reinvent the wheel, since a proliferation of different descriptive labels for similar linguistic phenomena creates confusion. Indeed, one of the challenges inherent in cross-linguistic comparison is identifying the differences, if any, between labels which may designate the same phenomenon, or which may imply a difference in perspective or in the actual form or meaning of a linguistic element. This is particularly apparent in the label 'discourse marker', but applies to all labels used for describing type noun usage.

---

[115] These issues are comprehensively presented by Haspelmath (2010), who argues that instead of syntactic categories, more universal concepts be selected which will be comparable across all languages.

A discussion of the descriptors coined for use in this thesis sheds light on the challenges involved in focusing on semantic-pragmatic function independent of its syntactic realization; these are the 'phoric modifier' and 'phoric similative' functions. As introduced in Chapter 7.2.3 and further elaborated in Chapter 7.3.3 and Chapter 8.3, the phoric modifier describes a communicative function in which the speaker draws attention to an individual object (or situation) whilst referencing a broader category of objects and comparing the object at hand to an element in the discourse. The discourse deictic aspect of the phoric modifier, designated by the descriptive label 'phoric', is the linguistic element most easily compared cross-linguistically, since anaphora and cataphora are high level concepts and discourse functions that can be understood independently of particular linguistic forms. In this respect, it is preferable to the syntactically based descriptive category 'postdeterminer/complex determiner construction' (De Smedt et al., 2007:227, referencing Denison, 2002) and versions thereof which have been adopted by all scholars since. Whereas 'postdeterminer' describes a particular syntactic word order and the oft-used 'postdeterminer complex' a type noun construction encompassing a phrase with the type noun in postdeterminer position, 'phoric' describes what the phrase accomplishes in discourse.

The same distance from syntax was not altogether achieved with the term 'modifier', which is rather grammatical in nature and was employed in order to differentiate the phoric modifier, which includes reference to a category, with the phoric similative, which does not. Although 'modifier' could be understood from a discursive, pragmatic perspective, the claim is tenuous at best. However, note that in order to qualify as a phoric modifier, the category reference need only be available in the discourse context and not necessarily present in the *type* phrase itself, i.e. it can be dropped. This makes phoric modifier as a functional category arguably more flexible than postdeterminer complex, examples of which in the literature tend to exhibit a fully binominal structure with no dropped elements. However, this may be because the languages studied do not permit dropped categories from binominal *type* constructions. (Here again we find ourselves more reliant on the syntactic structure than intended.) It seems that in certain cases, because of the nature of grammaticalization and in order to differentiate truly unique functions of type nouns from each other, it is necessary to employ some syntactic elements in descriptive labels, despite an attempt to make them purely semantic-pragmatic.

It has proven challenging to retain a focus on grammaticalization, a process which is inherently syntactic, and at the same time to define the semantic-pragmatic functions of *type* cognates without reference to syntax. To fully realize such a goal is perhaps a logical

contradiction. Yet the hope here is that by utilizing broader concepts such as 'phoric' and 'modifier', the process of cross-linguistic comparison across all language families can be more easily facilitated, leaving fewer languages excluded, if any. The fact that Semitic and Uralic languages have been successfully included in this study is perhaps evidence of some success in this regard. The same is true for the manner in which the comparison of the functions *type*, a noun, and *aisa*, a similative demonstrative, was facilitated. This leads to a second important element that has come to the fore during the course of this study, the importance of considering the whole lexicon while analysing particular lexemes, again focusing on function and not syntactic category.

10.2   *Type* grammaticalization in the context of the whole lexicon

There are many examples in the literature of studies comparing two type nouns of varying etymology, for example, Mihatsch (2016) which compares the binominal structures specifically of type nouns in Romance languages, expanding the list of type nouns beyond the common English three of *kind, sort* and *type* to include *class, form* and also *manner.* Later comparative studies have broadened the list of type nouns even further to include quasi type nouns such as *style, wave* and *cut* (Kornfeld, forthcoming). Although the majority of studies compare the grammaticalization of nouns exclusively, Mihatsch (2010) is an early example of research which focuses on the semantic-pragmatic function of hedging and compares how grammaticalized words from different syntactic categories, type nouns and similatives, are both employed as hedges. Voghera (2013) also draws attention to the complementary uses of Italian preposition *come* and (the originally nominal) grammaticalized *tipo* when functioning as similatives and equatives. The findings in both of these cases illustrate the value of focusing on semantic-pragmatic functions rather than syntactic categories.

Analysing lexemes in the context of the whole lexicon is of vital importance since two apparently similar lexemes can communicate subtly different semantic meaning and be used in slightly different pragmatic contexts. Examples of this were presented in Chapter 7.4.3 under the heading of functional specialization. Many studies have employed this approach, and in some cases scholars are able to provide reasons why lexemes have developed certain functions and not others, as the arguments in Chapter 7 demonstrated. In the Hindi data, analysing the functions of *type* in the context of other lexemes fulfilling the same functions has also provided insight into why *type* has not developed an exemplifier function, for example. Bar Chart 2 clearly illustrated the division of responsibility for fulfilling certain semantic-pragmatic functions among *aisa* and its related lexemes in particular, but also how *type* and other type nouns *prakaar* and *tarah* had developed specialisms within the lexicon.

Whilst certain semantic-pragmatic functions, such as hedge and quotative, were fulfilled by multiple lexemes, we can observe from Bar Chart 2 how adverbial *jaise* has a monopoly on the exemplification function; this provides a partial explanation for why *type* is never employed in this function.

The discovery of significant similarities between the innovative semantic-pragmatic functions of *type* and those of native Hindi *aisa* and its related lexemes led to a revision of the early conclusions of this thesis. Crucially, a comparison of the apparent grammaticalization of both *aisa* and *type* through the lens of De Smet et al.'s (2018) principle of attraction resulted in the native-lexeme-as-model hypothesis expounded in Chapter 9, namely, that a demonstrative construction with *type* is following the same path of grammaticalization as *aisa*. The way in which the noun *type* seems to have modelled itself on the grammaticalization process followed by the similative demonstrative *aisa* demonstrates that lexemes fulfilling the same function need not be from the same linguistic category. Because of this experience, the benefits of analysing *type* (or any other lexeme) in the context of other lexemes which fulfil the same semantic-pragmatic function regardless of their syntactic category cannot be overstated.

10.3 Innovative features of Hindi *type*

From one perspective, given the vast number of cross-linguistic similarities it exhibits, the development of *type* in Hindi is unremarkable, and yet there are a few factors that make it noteworthy: the relative speed of grammaticalization (within one or two generations) compared to *type* cognates in other languages, its status as a loanword still affecting its behaviour, and the development of a phoric similative function. We will review these in turn.

10.3.1 Speed of grammaticalization

In Chapter 3, I described the ways in which Hindi and Kumaoni speakers in Pithoragarh may have come into contact with English and therefore the word *type*, concluding that there was only limited opportunity for the lexeme to be borrowed prior to the advent of private English-medium education in the 1980s, after which *type* was most likely conventionalized and propagated, if not also initially borrowed. Even if *type* were in circulation in its categorizer function at the time of the census in 1921, that leaves fewer than 100 years prior to the recordings being made in 2015-16, which is a comparatively short time for a type noun to develop so many novel functions. This is in sharp contrast to how long it has taken in Romance languages (see Mihatsch (2016) for first attestation dates of taxonomic uses of *type* as well as other type nouns). In French, *type de* has not developed a well-established

approximative (hedging) usage, whereas other taxonomic nouns have (*genre de, espece, sorte de*). Mihatsch (2007:232) suggests this may be as a result of the late appearance of the taxonomic sense of *type* in French (in the eighteenth century), thereby hinting that a prolonged period is necessary for grammaticalization. The example in Fig 192 illustrates how an approximative use of *type* to refer to a Southern French dish similar to a pizza is not universally acceptable to native speakers:

> *? La pissaladière est un **type de** pizza.*
> 'The pissaladière is a **type of** pizza'
> **Fig 192: From Mihatsch (2007:232)**

The starkest contrast is with the original English word *type* itself, first employed as a categorizer in 1854, which has not grammaticalized beyond the ad hoc categorizer function. Yet both the English and French *type* cognates are relatively 'old' loans compared to Hindi *type*. Rosenkvist and Skärlund (2013) do not mention the date of first attestation of the taxonomic usage of Swedish *typ,* but their paper suggests that by 1930 the taxonomic meaning was already established. With anything from a 50- to 100-year disadvantage in comparison with its European cognates, the relative speed with which *type* seems to have developed so many highly grammaticalized functions in Hindi is notable. As Heine and Kuteva explain, 'grammaticalization is a gradual process that may involve generations of speakers and extend over centuries' (2005:81). Heine and Kuteva highlight the difficulties in making judgments on the causal effects of language contact on accelerated grammaticalization. Having described the Hindi lexicon in terms of a language-internal contact environment, with newcomer loanword *type* interacting with older loan *tarah* and Sanskrit-origin *aisa*, the same could apply to the native-lexeme-as-model hypothesis: 'It is more difficult to identify clear cases where language contact acted as an accelerating force in grammatical change, that is, where it was **causally responsible for speeding up what would have happened anyway**…' (2010: 95, **bold** mine). Heeding this word of caution, it is enough here to note the relative speed of development of *type* functions in Hindi versus the general cross-linguistic trend and to propose that the hypothesis that *type* is being attracted towards *aisa* and replicating its grammaticalization process is a plausible explanation for this rapid development.

10.3.2 Status as loanword still affects its behaviour

Backus (2005) argues that a speaker has the option of 'intentionality' or 'nonintentionality' when using a loanword as part of their discourse. The same speaker could use a word 'nonintentionally', i.e. as a default lexical item within their speech community, on one

occasion, and on another occasion use the word 'intentionally' to index association with the other language, i.e. as a single-word codeswitch: 'What distinguishes words of foreign origin from native words is that they can potentially be used to **exploit the indexical values associated with that language in the speech community**' (2005:319, **bold** mine). Whilst intentionality is hard to measure objectively, we know from discourse analysis and other disciplines that a speaker's language choices are meaningful and not random, and that there is (at least a degree of) agency on the part of the speaker who uses words in specific ways (i.e. intentionally) to achieve a communicative outcome. In Chapter 9.5.3 the use of *type* to flag English code-switches was explored. *Type* has indexical values associated with its origin as an English lexeme, which could explain why it is most frequently employed as a flag to alert the hearer to the presence another, less established, English word. Perhaps due to its recency of arrival in the lexicon, these indexical values may be present regardless of speaker intentionality since English code-switches could be flagged intentionally or subconsciously, as suggested earlier.

An alternative or additional explanation could be the phonotactics of the respective words, wherein an unknown (or lesser known) English-sounding lexeme triggers the phonotactically English flag *type* in preference to *jaisa*. Although each of the individual phonemes which constitute the Hindi pronunciation of *type* are commonly attested in Hindi words, their configuration as CDC, where D represents the diphthong /aɪ/, is not common. This may be a cause of the psycholinguistic priming which causes the speaker to preferentially employ *type* as a flag for certain English-origin words and not others, although further research is required to test this hypothesis.

10.3.3 Phoric similative function of *type*

Whilst many of the functions of English-origin loanword *type* in Pithoragarh Hindi mirror grammaticalized uses found in other languages, the phoric similative function of *type* has not been attested in other languages except partially in Norwegian, which demonstrates the apparent development of a *type* construction into a complex similative demonstrative, and where this *type* demonstrative construction is used as a phoric similative in some, but not all, possible contexts.

In Chapter 7.3 we observed how Norwegian *den type(n)* 'this/that type' had emerged as a chunk (fixed unit) which can be used in place of similative demonstrative *sånn* 'like this/that' in presentative constructions without a category noun following *type(n)*, e.g. *Jeg vet det finnes den typen* 'I know you get that type' (Odden (2019:180). However, whilst Norwegian

similative demonstrative *sånn* 'like this/that' can appear alone as a predicate, type noun constructions sound odd in this context, as Odden demonstrates with the more highly grammaticalized chunk *den slags* 'that kind' (Fig 193):

> *?De er **den slags***
> ?They are **that kind-GEN**

> *De er **sånne***
> They are **like that**
> **Fig 193: Copied from Odden (2019:161, example 66b; translation mine)**

In contrast, the equivalent Hindi construction *is type ka* 'like this' (lit. 'this type of') can and does appear as a predicate, as in Fig 114, reproduced here as Fig 194a. The utterance includes two similative demonstratives, simple *vaisi* and complex *us type ki*. However, it is notable that even without *vaisi*, the *type* phrase would be acceptable, as in Fig 194b.

> *Hai vaisi vo, **us type ki***
> She's like that, **like that** (lit. that type of)
> **Fig 194a - convo_d, Ankita (F28)**

> *Vo **us type ki** hai*
> She's **like that**
> **Fig 194b**

In the light of these findings, and the cross-linguistic comparison illustrated in Table 4, it seems Hindi loanword *type* is unique in forming, with the determiner, a similative demonstrative complex which can replace *aisa* in all phoric similative contexts. *Type* cognates in other languages seem to readily develop a similative meaning and function, and Romance languages plus English show complex phoric determiner constructions that foreground the deictic referent, leaving the subcategorization meaning in the background. However, none except Norwegian have come close to the pronominal and phoric similative uses shown by Hindi *type*.

Odden (2019:286) emphasizes the interchangeability between Norwegian demonstrative type noun construction *den type(n)* and similative demonstrative *sånn* and suggests that 'functional analogy' between the two may have resulted in the lexical chunk *den type(n)* acquiring some of the 'morphosyntactic properties' of *sånn*. Analogy is a key prerequisite in the principle of attraction expounded by De Smet et al. (2018). Odden highlights the use of *type* as a general extender, in fact the most common function of *type* in the Norwegian corpus studied, and concludes that *type* (and its more grammaticalized counterpart *slags* 'are [both] **widening their semantic-pragmatic scope** in line with the functions of *SÅNN*, and developing further polysemy' (**bold** mine). This finding in Norwegian, an unrelated language

with little or no contact with Hindi, is strikingly similar to that observed with *type* in Hindi and strengthens the argument proposed in this thesis that attraction between Hindi *type* and *aisa* has caused a widening of the semantic-pragmatic functions of Hindi *type*. Mihatsch (forthcoming a) notes that complex postdeterminers are a source for general extenders cross-linguistically, so it is possible that the phrase *is/us type ka* would have developed into a general extender anyway. Yet, as discussed in Chapter 7.2.5, these usually involve a combination with a vague noun, which we do not see in the Hindi data. Although this function of *type* in Hindi, apparently replicating that of *aisa,* is unique among languages studied, it is relevant to remember that as described in Chapter 7.4, there is a cross-linguistic tendency for *type* to interact closely with similative demonstratives. Additionally, there seems to be a special relationship between *type* and demonstratives, as Mihatsch explains (2016:152): 'Nouns equivalent to *type* are most frequent in the phoric function… Determiners preceding the type noun are demonstratives – Italian *quel* and *questo* Portuguese *esse*, *este* and *aquele*, Spanish *ese* and *este*… and the definite article.' Possibly due to the hugely prolific role of *aisa* and its related lexemes in colloquial Hindi, which this thesis has uncovered, and certainly aided by the area of overlap in the phoric modifier function, it seems the attraction between *aisa* and *type* has been strong enough to trigger the grammaticalization of *type* towards functioning as a phoric similative. That this is not in evidence in the web data is one of a small number of interesting differences between the two data sets which will be discussed in the following section.

10.4   Differences between informal spoken and written data sets

The brief review of the Hindi Web data in Chapter 8.7 led to an important discovery, that there are clear differences in the ways that Hindi speakers use loanword *type* in written compared to spoken discourse, even when both are of an informal nature. Both semantic-pragmatic and syntactic differences have been identified. Although many of the semantic-pragmatic functions of *type* from the spoken data were attested in the written data, there were notably no examples of the phoric similative function or other extended uses of the *type* demonstrative construction *is type ka*. Conversely, there was a function of *type* as a person descriptor that hardly features in the Pithoragarh corpus but is prolific in the web data. From a grammatical perspective, we observed *type* appearing alone without the genitive and even as an utterance-final tag modifying phrases and clauses.

With the exception of Janebová et al. (forthcoming), the research on type noun grammaticalization conducted thus far has not focused on a comparison of written versus spoken media, although it is notable that the highly innovative uses of *type* cognates cross-linguistically (and indeed of similative demonstratives) are usually described in the context of youth language.[116] In what follows, the possible reasons underlying the differences in findings between the spoken and written data analysed in this study are outlined and discussed.

10.4.1 Web data shows more highly grammaticalized uses of *type*

The brief discussion afforded within Chapter 8.7 itself indicates that the aspects of diversion from the spoken data set are still within the range of variation exhibited by cross-linguistic data. The use of bare *type* (with no genitive) was attested in the spoken corpus, but only in a few instances (6% of tokens in the spoken corpus compared to 19% in the web corpus). Additionally, in the majority of cases in the web data, bare *type* marked a complete clause, especially (quasi-)quoted text. The progression of scope from marking a single word or phrase to a whole clause is evidence of further grammaticalization. We note from other languages that from a syntax perspective *type* develops along a cline from a noun to a preposition, to an adverb with phrasal scope and then an adverb with sentential scope, and finally a discourse particle, which is independent of sentence structure altogether. Russian does not attest adverbial usage and the most grammaticalized uses of *tipa* are analysed as discourse particle usage (Kolyaseva and Davidse, 2018); however, for languages in which all syntactic categories are attested, the order from least to most grammaticalized progresses as described. This means that in line with the cross-linguistic precedent, we can interpret the Hindi web data, with its higher percentage of sentential scope examples, as exhibiting more highly grammaticalized uses of *type*.

The Hindi produced by bloggers and those who comment on them is likely to exhibit more language change because they are more likely to be writing from bigger cities and because the language spoken in bigger cities is more likely to include innovative language use, which in this context means 'more grammaticalized' (see Hurst (2017) on the rural/urban divide). In addition to the urban/rural divide we can also mention age, income, and education level as factors that influence the likelihood of someone to not only be using the web but actively engaging in web-based forums and writing blogs. We may also expect those participating in

---

[116] See, among others, Kolyaseva (2018) and Mihatsch (2018) for type nouns; and Ekberg, Opsahl and Wiese (2015), and Schoning, Helder and Diskin-Holdaway (2021) for similative demonstratives.

these activities to have a higher degree of bilingualism, which would also make them more likely to be linguistic innovators. Additionally, we know that grammatical change develops first in spoken language and is later used in written form (Heine and Kuteva, 2005:250), so we can assume that the Hindi spoken by the bloggers and comment writers is at least as grammaticalized as that which they use in their writing. For these reasons, the Hindi Web data can be considered further along the grammaticalization path than the spoken data from Pithoragarh.

There are cognitive reasons to support the theory that the grammaticalized construction is/us *type ka* 'like this/that' could be shortened to *type*. Elision usually occurs from the end of a word or phrase; since in grammaticalized uses of *is type ka*, i.e. not as a phoric modifier, the genitive *ka* is the final syllable of the phrase, it is most likely to be omitted. In addition, we have seen how the genitive *ka* seems to have become (or are becoming) cliticized and there are cases of it being elided in a hedging context (see Chapter 8.5.2). It is likely on the basis of cross-linguistic evidence that "N/ADJ + *TYPE* + GEN" could grammaticalize further to bare *type* as the Hindi Web data suggests is already happening online. In fact, an interesting increase in colloquial usage of *type* has been perceived by the researcher in recent years (since the original data were collected). Notably, in the past 18 months during informal conversations in Hindi via WhatsApp, one friend (not a research participant) has used *type* extensively, much more frequently than the researcher herself and in ways never knowingly used by the researcher, including as a hedge modifying a complete utterance (Fig 195) and as a quotative (in both cases using bare *type* without the genitive). Other uses by the same friend during the previous six months include *type* functioning as a categorizer, a phoric modifier, a phoric similative (both with and without the genitive: *is type ka* and *is type*), a hedge with the genitive, and more frequently a hedge without the genitive, as in Fig 195*.* Note that most people choose to message in "Roman Hindi", as the example illustrates.

> Friend: *Aakash me bahut dur.....kuch hua….orange color sa. Thoda sa….jese koi dhamaka hua ho....*
> Far away in the sky…..something happened….kind of an orange colour. A bit….like some explosion might have occurred….
>
> Self: *Dhamaka matlab*
> *Dhamaka* meaning…?
>
> Friend: *Blast **type***
> **Kind of** a blast
>
> Self: *Hmm*
> Hmm

Friend: *Pr blast ni tha*
But it wasn't a blast

The number of tokens in which *type* appears without the genitive is noteworthy. In the example in Fig 196, *type* is used to hedge the adjective *ajeeb* 'strange'. The hedge appears within a phrase *ajeeb wali English* 'strange kind of English' (lit. 'strange one English'), which creates an odd effect grammatically, but is perfectly understandable from a communicative perspective. The text is reproduced exactly as it was in WhatsApp, in three separate lines. The second use of *type* as a hedge, also in its bare form without the genitive, occurs in the third line and acts as a hedge on the complete proposition stated in the previous two lines.

ajeeb **type** *wali English lg ri na meri*
My English seems a **kind of** strange one, doesn't it?

jese bna bna k bol ri hu
As if I'm making it up as I go along

Type
Kind of thing

Fig 196: Friend, 11ᵗʰ November 2021; text reproduced exactly as in WhatsApp

This anecdotal data may be a reflection of an ongoing trend towards greater grammaticalization, characterised by both the elision of the genitive and by the tendency for *type* to have scope over whole utterances. It is certainly a development that we would expect to see from a cross-linguistic perspective, and also seems to be predicted from the Hindi Web data.

10.4.2 Lack of phoric functions of *type* in the web data

Informal conversational discourse is dynamic and interactional. One of the strategies utilized in real-time speech is anaphora, which allows speaker and hearer to refer to objects either in their shared physical space, or available to both via previous discourse or shared knowledge. We would expect anaphora to be employed to a much lesser extent when communicating via written media such as a blog post, which is less interactional and more static, although anaphoric reference could appear in comment sections and online forums where one author responds directly to what another has written. Demonstratives (this/that) are a key linguistic tool for phoric reference, and it is therefore unsurprising that of the two kinds of Hindi data, the spoken corpus showed a higher percentage of phoric uses of the demonstrative with *type* (24%) compared to the written data (only 7%). Janebová et al. (forthcoming) note that uses of similative demonstrative *takový* 'such' in Czech are five times

higher in spoken than in written data and are predominantly recognitional (non-phoric), creating a connection between speaker and hearer based on shared knowledge. However, taking a closer look at the data, we observe that both corpora contain categorizer and phoric modifier uses of *type*: the principal difference between the two is the complete absence of the phoric similative function in the written data. This could be due to the same reason under discussion here, since the vaguer the phoric reference, the harder it is to resolve, and perhaps only real-time interlocutors (who know each other well) could understand each other's phoric similative uses. The other possible explanation for this phenomenon is that the dialect of Hindi developed in Pithoragarh (perhaps in the Kumaon region as a whole) is unique in developing the phoric similative function.

Furthermore, the preference for using anaphora as a discourse mechanism does not explain the innovative uses of the *type* phrase *is type ka*[117] as a general extender, quotative or filler. However, general extenders and especially fillers are also more likely to occur in spontaneous discourse where the speaker does not have as much time for retrieving exact words. As far as quotatives are concerned, the web data employs the single bare *type* which, as mentioned above, is a more grammaticalized form. However, it would be tenuous to argue that the authors had first employed the demonstrative phrase *is type ka* and later dropped the demonstrative and the genitive, since we have no evidence either for against this hypothesis. Indeed, the possibility that Hindi *type* is grammaticalizing differently in different dialects and regions of North India must be considered, but without the location and sociolinguistic profile of the web data authors, it cannot be tackled within the current data sets alone. More extensive data collection among Hindi speakers from different parts of North India is required in order to investigate how *type* usage compares across dialects and population groups.

10.4.3 Expressive descriptions of people

One discourse function of *type* which was attested in both spoken and written corpora was its use in descriptions of people. There is a distinct lack of attributive modifier and ad hoc categorizer uses of *type* in the spoken data from Pithoragarh. By contrast, the written data showed a substantial number of modifier uses akin to the English attributive modifier or semi-suffix, which allow the speaker to describe something (or more commonly, someone) in unusual, often evaluative, terms. Examples include more ordinary adjectives (*anpadh*

---

[117] All five of the extended uses of the complex similative demonstrative with *type* use the proximal demonstrative *is* 'this' and not the distal demonstrative.

'uneducated'; *buddhijevi* (hic) 'intellectual') as well as highly colourful quasi-adjectives such as that in Fig 197. Many of the descriptions involve similative comparison with a named individual (see the example of the politician *Diggi* in Fig 182).

कुछ कुछ **इन्नोवेटिव टाइप के परेशान होने वाले** परेशानियों का वज्रगुणन कर लेते हैं- घर में दफ़्तर से परेशान होते हैं और दफ़्तर में घर से।

*kuch kuch **innovative type ke pareshaan hone waale** pareshaaniyon ka vajraagunan kar lete hain- ghar me daftar se pareshaan hote hain aur daftar me ghar se.*

Some **innovative kinds of stressed people** multiply their worries big time - at home they stress about work, and at work they stress about home.

Fig 197: From http://fursatiya.blogspot.com/2009/08/

In comparison to the elaborate expressive descriptions, we can observe a preference in the Pithoragarh corpus for single adjectives hedged with *type ka* (*type* and a genitive). One reason for the difference between these two strategies employed across the corpora could be the additional time afforded a blogger to formulate more complex and imaginative comparisons or innovative descriptions. Informal conversational discourse is fast-paced compared to the potentially unlimited amount of time available to bloggers for composing the written word. In real-time speech, it takes more cognitive effort to invent eloquent descriptions; indeed, it is possible that the same bloggers would resort to use of simple nouns and adjectives to describe someone in their speech. The Pithoragarh data also showed a slight preference for hedging (N=26) over similative comparison (N=24). This could also be due to the shorter formulation of the hedge, e.g. ADJ + JAISA, but is also due to a preference for Kumaoni descriptive adjectives and nouns (10/26).

10.5   Conclusion

It will have become clear through the topics discussed in this chapter that the development of English loanword *type* in colloquial Hindi as a phoric similative function is unique cross-linguistically and it appears that the complex similative demonstrative phrase *is type ka* is grammaticalizing in ways not observable in the web data. Despite the challenges of cross-linguistic comparison, the benefits are unparalleled, particularly in situations where diachronic data is unavailable but the cross-linguistic tendencies are so universal, as with type noun grammaticalization. A comparison between the spoken and written data revealed several differences which can be explained by the differing nature of interaction, although more research is required in order to determine whether the innovative functions of *type* attested in the spoken data from Pithoragarh are dialect specific or replicated across other Hindi-speaking communities.

Opportunities for furthering the research begun here will be presented in the final chapter, along with a review of the contributions and implications of this study.

11   Conclusion

This study of spoken colloquial Hindi from a remote community offers a unique insight into language use away from India's urban centres, perhaps more reflective of the Hindi spoken in the majority of North India, which, like Pithoragarh, is (semi-)rural. The naturalistic recordings contain evidence of *type* usage in spoken language from the past decade and permit us to analyse the functions of a type noun which seems to be in the middle of a relatively speedy process of grammaticalization. This offers a fairly unique opportunity compared to other languages studied (see Chapter 7), in which the process of grammaticalization has been continuing over several centuries and may in some cases be largely completed. The comparison between spoken and written (web) data has illustrated that type nouns are employed differently in written text, where the time afforded the author allows for more expressive descriptions resulting in novel uses of *type*, e.g. as a person descriptor.

The cross-linguistic analysis conducted in Chapter 7 is the first of its kind to focus solely on the lexeme *type* and its cognates across multiple languages. Although this approach loses some of the richness a comparison with other type nouns would have afforded, it gains in having a distilled focus which allows us to concentrate on issues pertaining specifically to *type*. Indeed, the single-minded spotlight on *type* has brought into focus the cross-linguistic tendency for *type* cognates to interact with similative demonstratives (see Chapter 7.4). This sustained focus on both classes of lexemes and the interaction between them from the perspective of their shared semantic-pragmatic functions is also unique, and combines two recent and productive areas of grammaticalization research.

The study has implications for research into the grammaticalization of type nouns more generally. Firstly, we see that analysis of equivalent lexemes is essential. Type nouns do not operate alone but share semantic-pragmatic functions with other (partially) synonymous items in the lexicon. Adapting the Firthian maxim, "You shall know a word by the company it keeps", it is clear that in order to fully understand the meaning of a word, scholars must look at the alternatives at the speaker's disposal and analyse the differences in usage between them. A wealth of research in the volume by Brems et al. (forthcoming) indeed follows this approach.

Secondly, the notion of 'contact' can be extended to language-internal environments and processes. It has been argued in this thesis that the relationship between *type* and its equivalent lexeme *aisa* is at once language-internal and also symptomatic of a language

contact environment where 'native' and 'foreign' lexemes meet. The enduring dual identity of *type* as an English-origin word in addition to its status in the Hindi lexicon is one example of this. Additionally, the native-lexeme-as-model theory has the potential to explain how loanwords more generally can grammaticalize through analogical replication encouraged by attraction to a native lexeme.

Owing to the wide array of areas impacted by this study, there are many opportunities for progressing the work further. In this section, suggestions for future research in the following areas will be highlighted: further analysis of *type* synonyms in the current data set; diachronic data to confirm and develop the grammaticalization hypotheses; comparison with functions of *type* in other dialects of spoken Hindi throughout North India; and (further) analysis of both *type* and *aisa* and their synonyms in the Hindi Web corpus and in social media.

We have seen how the functions of *type* in colloquial Hindi discussed in Chapter 8 are fulfilled not only by *type* but even more so by *aisa* and its related lexemes and, in the case of general extenders, quotatives and discourse markers, by other lexemes in the Hindi lexicon. Each one of these semantic-pragmatic functions is worthy of detailed study. For example, the spoken data set is rich in new quotatives, including *aisa* and *type* but also other strategies mentioned in Chapter 8.6.3. There is much to be done to continue the research begun here on the similative demonstrative paradigm in Hindi. Due to the highly multifunctional nature of *aisa* and its related lexemes, it has not been possible to investigate each aspect of its many uses, and it is therefore hoped that this thesis may serve as a launchpad into more granular studies. To mention but one possibility, the discourse and narrative functions of *aisa* and *vaise*, including their roles in structuring (whole stretches of) discourse, are worthy of detailed analysis. The comparison between *type* and its equivalent lexemes begun in Chapter 9 illustrates the great scope for further research using the spoken data collected for this thesis. Similarly, retrieving tokens of *aisa* (and *jaisa*) usage from the Hindi Web corpus would yield more productive comparisons.

From a sociolinguistic point of view, the current data set would allow a focus on such sociolinguistic variables as topic of conversation and gender. Although it was not possible to use age grading due to the limited number of *type* tokens, the opportunities for using the current data set to analyse how *aisa* and related lexemes are employed by speakers of different ages and genders are more extensive. These are only a few of the many areas of further interest which have been generated by this study and for which there are varied opportunities for continued research. Particular emphasis should be placed on the use of

lexemes in their functional (semantic-pragmatic) context, and on emerging uses which are not yet dictionary-attested.

The previous chapter highlighted some limitations of the current synchronic data set. In terms of future research, it will be conceivably impossible to source historical spoken data from Pithoragarh, and informal written data such as the Hindi Web data discussed in Chapters 8.7 and 10.4 can offer a useful comparison but do not aid our understanding of the grammatical development of demonstrative constructions with *type* which were found to be underrepresented in written language. If similar research were to be conducted in other Hindi-speaking areas of North India, a comparison between the results may highlight nuances related to the Pithoragarh dialect. Additionally, a longitudinal study of *type* usage in Pithoragarh akin to that conducted by Gardner-Chloros (2013) on code-switching patterns in Strasbourg, or prior to that Trudgill (1988) on urban dialect in Norwich, would offer insight into the ongoing process of grammaticalization.
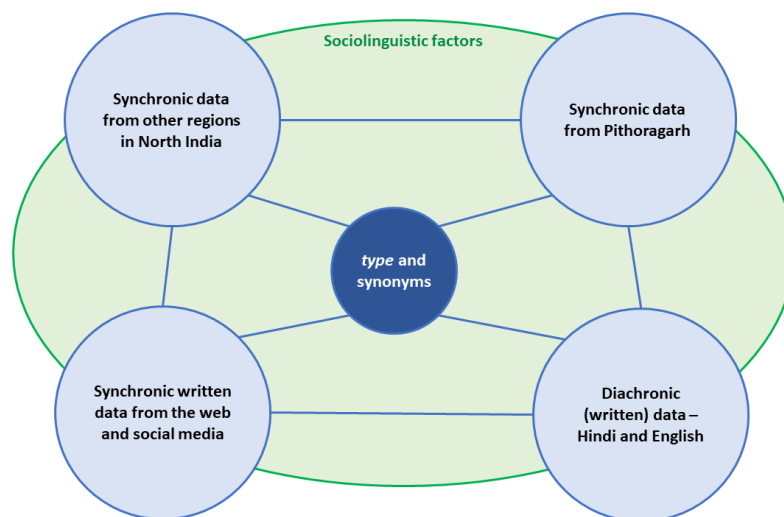
Fig 198: Future research opportunities within Hindi

The diagram in Fig 198 depicts the potential synergy between different data sets that could be sourced (where needed), analysed and compared with one another. Each data set offers a new perspective on the uses of *type* and its synonyms and the grammaticalization process(es) in which they have partaken, and each can be compared with each of the others to yield useful insights. Chapter 10.4.1 included brief references to WhatsApp conversations: there is enormous scope for the analysis of informal chats from this social medium due to its prevalence in daily life in North India. Analysis of sociolinguistic variables associated with each data set would shed light on the use of different lexemes in different contexts by different speakers. For example, considering the age and gender of the speaker/author as

well as their location (rural/urban) and their other languages would add several different dimensions to our understanding of which factors shape language choices regarding *type*, *aisa* and their related lexemes.

It will have become clear from the above and throughout this thesis that the work presented herein is at once innovative and yet continues in the vein of increasingly detailed and insightful research on the grammaticalization of type nouns and similative demonstratives conducted by colleagues around the world. Investigation into the English loanword *type* has presented an opportunity to describe and analyse the prolific but largely overlooked similative demonstrative *aisa* and its related lexemes. From the list of possibilities for future research proposed here it will be clear that there is much work still to be done.

APPENDIX A: Permission Form

भाषा वैज्ञानिक शोध – पिथौरागढ़ के रहने वालों की हिन्दी

पी.ऐच.डी की शोधक:- आराधना[118] वाल्टज़

यह शोध पिथौरागढ़ के रहने वालों की हिन्दी के ऊपर है। आप किसी और के साथ बात करेंगे और आपके शब्द रिकोर्डर में रिकोर्ड हो जायेंगे और इसलिये कि आपकी बोली हुई बातें हमें अच्छी तरह समझ में आ जाये आपका विडियो भी बनाया जायेगा।

हम जानते हैं कि हिन्दी एक भाषा है लेकिन लोगों द्वारा इसे अलग-अलग तरीके से बोला जाता है। इस शोध में मैं जानना चाहती हूँ कि लोगों की हिन्दी बोलने में क्या अंतर है।

आपकी बोली हुई बातें कम्प्यूटर में रखी जायेंगी और सख्त गोपनीय रखी जायेंगीं। रिकोर्डिंग करने के बाद सारे सहभागियों की भाषा को विश्लेषण किया जायेगा और उस विश्लेषण का परिणाम मेरी पी.ऐच.डी के निबंध में लिखा जायेगा। आपका नाम गुप्त रखा जायेगा। इस शोध में मैं आपका सहयोग चाहती हूँ।

....

शपथ:-

मुझे इस शोध के बारे में बताया गया है और मैं भाग लेने के लिये तैयार हूँ।

मैं जो बोलूँगा उसका प्रयोग शोध के लिये और शिक्षा के लिये किया जा सकता है।

मैं समझता हूँ कि अगर मैं बीच में भाग ना लेने का फैसला करूँ तो मेरे द्वारा जितना बोला गया वह नष्ट किया जायेगा।

मैं यह भी समझता हूँ कि मेरी बोली हुई बातें सख्त गोपनीय रखी जायेंगीं।

मेरी आयू 16 साल से ऊपर है।

---

[118] Aradhana is the researcher's local Hindi name.

नाम _____

228

हस्ताक्षर _____

आज की तारीख _____

इस शोध में भाग लेने के लिये धन्यवाद!

APPENDIX B: Features of Pithoragarh Hindi

The following is a list of dialectal features which are common to the Hindi spoken by residents of Pithoragarh and would not be considered part of standard Hindi. I have compiled the list based on 15 years of participant observation as a Hindi-speaking member of the community and corroborated in the recorded data from the Pithoragarh corpus. The list is not exclusive, but contains many of the features of the dialect which will become apparent from reading the examples in the thesis.

1. **Extensive use of progressive aspect**
2. **Code-switching with Kumaoni**
3. **Established loanwords from Kumaoni**

   Nouns (often for specifically local objects)

   Adjectives (usually for describing people)

   Discourse markers: *bhal*, *dhen, oija*
4. **Use of negative marker for soliciting agreement with interlocutor**
5. **Omission of perfect aspect auxiliary**
6. **No distinct feminine plural verbal conjugation** (masculine plural conjugation is used for both)
7. **Use of 3SG pronoun in place of plural pronoun**

APPENDIX C: Summary of recordings and participants by group

Recording group	No. of speakers	Description of group	No. of audio files	Total length
convo_a	5	Family members across two generations	1	01:06:28
convo_b*	5**	Neighbours living in the same building owned by one of the participants (landlady)	1	01:30:56
convo_c	Removed from sample			
convo_d	2	Sisters-in-law living in joint family home	1^	01:00:20
convo_e	3	Students at ILC as well as at local university (one of them also an ILC teacher for lower level classes)	1	01:00:00
convo_f	2-4**	Friends (unmarried ladies), joined briefly by a friend and relative	3	01:00:47
convo_g	3	Middle-aged relatives (women)	1	01:01:59
convo_h*	2-5	Friends (boys), later joined by family members across two generations	3	01:09:13
convo_i	3	Neighbours (women) - conversation mostly in Kumaoni	2	01:05:18
convo_j	2	Friends (men), who took the digital recorder with them on a Sunday drive through the hills	2	00:49:54
convo_k	3	Husband and wife, joined briefly by a young neighbour	1	01:27:31
convo_m	3	Cousins (three men, one of them an ILC teacher)	1	01:02:30
convo_n	3	Final-year school students (girls)	1	01:04:20
convo_p	3	Sisters-in-law living in joint family home, joined briefly by a neighbour	2	00:53:25
convo_q	2	Mother and (adult) son	1	01:01:10
convo_t	3	Husband and wife, and their (male) friend	1	01:10:37
TOTAL	**47****	-	**22**	**16:24:33**

* Researcher present for the final 20 minutes of the recording. In convo_b the researcher was actively engaged in conversation, whereas in convo_h she was sitting quietly alongside the family, people who she knew very well.

** Anita was present in both convo_a and convo_b, and Mohan was present in both convo_f and convo_j, but they have only been counted once in the total.

^ There was a second recording of only 3 seconds whose single utterance was moved to the main file for convenience

भाषा वैज्ञानिक शोध – पिथौरागढ़ के रहने वालों की हिन्दी

कृप्या आप इमानदारी से अपने उतर दीजिये।

नाम..

मोबाइल नंबर...

D.O.B..

निवास स्थान..

आप क्या करते हैं? ....

आप किसके साथ रहते हैं? ....

निम्नलिखित प्रश्न आपकी शिक्षा और रहने के बारे में पूछे जा रहे हैं।

आपकी शिक्षा किस भाषा के द्वारा की गयी है?

	क 1 - 5	क 6 - 8	क 9 - 12	BA/BCom/ BSc/BTech	MA/MCom / MSc/MTech	PhD
हिन्दी						
अंग्रेज़ी						
अन्य भाषा						

आप अपनी ज़िन्दगी में कहाँ-कहाँ रह चुके हैं? 18 साल के बाद भी भर दीजिये।

	गाँव में	पिथौरागढ़ शहर में	उत्तराखण्ड के अन्य शहर में	अन्य जगह
आयूँ 0-5				
आ. 5-10				
आ. 10-14				
आ. 14-18				

क्या आपने कभी इंग्लिश स्पीकिंग की क्लास ली? कितने महिने?

निम्नलिखित प्रश्न आपकी दैनिक-दिनचर्या की भाषा के बारे में पूछे जा रहे हैं।

आप अपनी माँ से किस भाषा में बात करते हैं?

	हमेशा	अधिकतर	कभी-कभी	विषेश अवसर	कभी नहीं
पहाड़ी					
हिन्दी					
अंग्रेज़ी					
अन्य भाषा					

आप अपने पिता-जी से किस भाषा में बात करते हैं?

	हमेशा	अधिकतर	कभी-कभी	विषेश अवसर	कभी नहीं
पहाड़ी					
हिन्दी					
अंग्रेज़ी					
अन्य भाषा					

आप अपने भाई-बहनों से किस भाषा में बात करते हैं?

	हमेशा	अधिकतर	कभी-कभी	विषेश अवसर	कभी नहीं
पहाड़ी					
हिन्दी					
अंग्रेज़ी					
अन्य भाषा					

आप अपने दादा-दादी/बच्चों से किस भाषा में बात करते हैं?

	हमेशा	अधिकतर	कभी-कभी	विषेश अवसर	कभी नहीं
पहाड़ी					
हिन्दी					
अंग्रेज़ी					
अन्य भाषा					

आप अपने रिश्तेदारों से किस भाषा में बात करते हैं?

	हमेशा	अधिकतर	कभी-कभी	विषेश अवसर	कभी नहीं
पहाड़ी					
हिन्दी					
अंग्रेज़ी					
अन्य भाषा					

आप अपने दोस्तों से किस भाषा में बात करते हैं?

	हमेशा	अधिकतर	कभी-कभी	विषेश अवसर	कभी नहीं
पहाड़ी					
हिन्दी					
अंग्रेज़ी					
अन्य भाषा					

आप उँच्छ वर्गीय कर्मचारियों से किस भाषा में बात करते हैं?

	हमेशा	अधिकतर	कभी-कभी	विषेश अवसर	कभी नहीं
पहाड़ी					
हिन्दी					
अंग्रेज़ी					
अन्य भाषा					

आप अपने से नीचे कर्मचारियों से किस भाषा में बात करते हैं?

	हमेशा	अधिकतर	कभी-कभी	विषेश अवसर	कभी नहीं
पहाड़ी					
हिन्दी					
अंग्रेज़ी					
अन्य भाषा					

आप अपने समवर्गीय कर्मचारियों से किस भाषा में बात करते हैं?

	हमेशा	अधिकतर	कभी-कभी	विषेश अवसर	कभी नहीं
पहाड़ी					
हिन्दी					
अंग्रेज़ी					
अन्य भाषा					

निम्नलिखित प्रश्न आपकी अंग्रेज़ी भाषा के बारे में पूछे जा रहे हैं।

आपको कितनी अंग्रेज़ी आती है?

- ❖ मैं बिलकुल नहीं जानता/जानती हूँ।
- ❖ मैं कुछ शब्द जानता/जानती हूँ।
- ❖ मैं थोड़ा बहुत बोल सकता/सकती हूँ।
- ❖ मैं बहुत लिख सकता/सकती हूँ पर कम बोल सकता/सकती हूँ।
- ❖ मैं बोल सकता/सकती हूँ पर अपनी पूरी बात नहीं रख सकता/सकती हूँ।
- ❖ मैं बहुत अच्छी तरह बोल सकता/सकती हूँ।

आप अंग्रेज़ी भाषा किस-किस स्थिति में प्रयोग करते हैं? जितने सही हैं, सब में निशान लगायें।

- ❖ अखबार पढ़ने में
- ❖ टी.वी. देखने में
- ❖ फिल्म देखने में
- ❖ परिवार के साथ बात करने में
- ❖ दोस्तों के साथ बात करने में
- ❖ अपनी नौकरी करने में
- ❖ किताब पढ़ने में
- ❖ कम्प्यूटर चलाने में
- ❖ वट्स एप, एस.एम.एस. इत्यादि करने में
- ❖ फेसबुक चलाने में
- ❖ अन्य प्रयोग..

निम्नलिखित प्रश्न आपके विचारों के बारे में पूछे जा रहे हैं।

आप निम्नलिखित वाक्यों से कितना सहमत हैं?

1= बिलकुल सहमत नहीं हूँ   7=पूर्ण रूप से सहमत हूँ

1 और 7 के बीच अपने विचार के हिसाब से किसी भी नंबर में निशान लगाना।

	1	2	3	4	5	6	7
1. मुझे अंग्रेज़ी अच्छी नहीं लगती।							
2. मुझे अंग्रेज़ी बोलने का शौक है।							
3. मेरे भविष्य के लिये अंग्रेज़ी बहुत ज़रूरी है।							
4. मुझे जितनी अंग्रेज़ी आती है, उतनी काफ़ी है।							
5. अपनी जगह अंग्रेज़ी ठीक है पर उसे हिन्दी से अधिक महत्व नहीं रखनी चहिये।							
6. भारतीय सरकार अंग्रेज़ी को बढ़ावा देती है।							
7. मैंने मजबूरी से अंग्रेज़ी पढ़ी।							
8. जिन लोगों को ज़्यादा अंग्रेज़ी आती है, वे ज़्यादा पढ़े-लिखे होते हैं।							
9. मेरे उम्र के लोगों को अंग्रेज़ी की ज़रूरत नहीं होती है।							
10. अंग्रेज़ी सीखने से लोगों का आत्मविश्वास बढ़ जाता है।							
11. जब मैं बाहर जाता/जाती हूँ तो मुझे अंग्रेज़ी की ज़रूरत होती है।							
12. जिन लोगों को ज़्यादा अंग्रेज़ी आती है, उनको ज़्यादा आदर मिलता है।							

References

Texts

Aaron, J.E., 2016. The road already traveled: Constructional analogy in lexico-syntactic change. *Studies in Language 40*; Amsterdam, 26–62.

Aijmer, K., 2002. *English Discourse Particles: Evidence from a Corpus*. John Benjamins Publishing.

Aijmer, K., 2013. *Understanding Pragmatic Markers.* Edinburgh University Press.

Aijmer, K., forthcoming. Czech type nouns: Evidence from corpora, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move*. De Gruyter.

Andersen, G., 2010. A contrastive approach to vague nouns, in: Kaltenböck, G., Mihatsch, W., Schneider, S. (Eds.), *New Approaches to Hedging*, Studies in Pragmatics. Emerald, Bingley, UK, 35–48.

Anderson-Finch, S., 2011. 'More than the sum of its parts: Hinglish as an additional communicative resource.' In Kothari, R., Snell, R. (Eds.). *Chutnefying English: the phenomenon of Hinglish*. Penguin Books, New Delhi, 53-70.

Annamalai, E., 2001. *Managing multilingualism in India: political and linguistic manifestations.* SAGE, London.

Backus, A., 2005. Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism 9*, 307–340.

Badley, B.T., 1931. *Visions and Victories in Hindustan: A Story of the Mission Stations of the Methodist Episcopal Church in Southern Asia*. Methodist Publishing House.

Bhatia, Tej, K., 2011. 'The multilingual mind, optimization theory, and Hinglish'. In Kothari, R., Snell, R. (Eds.) *Chutnefying English: the phenomenon of Hinglish*. Penguin Books, New Delhi, 37-52.

Bittencourt, V.O., 1999. Gramaticalização e discusivização no português oral do Brasil: o caso "tipo (assim)." *Scripta 3*, 39–53.

Borowiak, T., 2012. English-induced changes to urban colloquial Hindi lexicon and structure. Mixed speech as a mode of discourse. *Lingua Posnaniensis 54*, 35–44.

Breban, T., 2010. *English Adjectives of Comparison: Lexical and Grammaticalized Uses.* Walter de Gruyter.

Brems, L., 2011. *Layering of Size and Type Noun Constructions in English*. Walter de Gruyter, Berlin; Boston.

Brems, L., Davidse, K., 2010. The Grammaticalisation of Nominal Type Noun Constructions with kind/sort of: Chronology and Paths of Change. *English Studies 91*, 180–202. https://doi.org/10.1080/00138380903355023

Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W., forthcoming. *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move.* De Gruyter.

Brown, P., Levinson, S.C., 1987. *Politeness: some universals in language usage*, Studies in interactional sociolinguistics. Cambridge University Press, Cambridge.

Buchstaller, I., 2013 (2014). *Quotatives: New Trends and Sociolinguistic Implications*. John Wiley & Sons.

Buchstaller, I., van Alphen, I. (Eds.), 2012. *Quotatives: cross-linguistic and cross-disciplinary perspectives*. John Benjamins Pub. Co, Amsterdam; Philadelphia.

Chand, V., 2011. Elite positionings towards Hindi: Language policies, political stances and language competence in India. *Journal of Sociolinguistics 15*, 6–35.

Cheshire, J., 2007. Discourse variation, grammaticalisation and stuff like that. *Journal of Sociolinguistics 11*, 155–193. https://doi.org/10.1111/j.1467-9841.2007.00317.x

Cruse, D.A., 1986. *Lexical Semantics*. Cambridge University Press.

Davidse, K., Brems, L., De Smedt, L., 2008. Type noun uses in the English NP: A case of right to left layering. *International Journal of Corpus Linguistics 13*, 139–168.

De Smedt, L., Brems, L., Davidse, K., 2007. NP-internal functions and extended uses of the 'type' nouns kind, sort, and type: towards a comprehensive, corpus-based description, in: Facchinetti, R. (Ed.), *Corpus Linguistics 25 Years On*. Rodopi, Amsterdam, 225–255.

De Smet, H., D'hoedt, F., Fonteyn, L., Van Goethem, K., 2018. The changing functions of competing forms: Attraction and differentiation. *Cognitive Linguistics 29*, 197–234.

Diessel, H., 1999. *Demonstratives: Form, function and grammaticalization*. John Benjamins Publishing.

Diessel, H., 2006. Demonstratives, joint attention, and the emergence of grammar. *Cognitive Linguistics 17*, 463–489. https://doi.org/10.1515/COG.2006.015

Diewald, G., 2011. Pragmaticalization (defined) as grammaticalization of discourse functions. *Linguistics 49*, 365–390.

Durkin, P., 2014. *Borrowed words: a history of loanwords in English.* Oxford University Press, New York.

Edwards, M., Gardner-Chloros, P., 2007. Compound verbs in codeswitching: Bilinguals making do? *International Journal of Bilingualism 11*, 73–91.

Ekberg, L., 2010. Extended uses of sån ('such') among adolescents in multilingual Malmö. *Young urban Swedish: Variation and change in multilingual settings,* 49–65.

Ekberg, L., Opsahl, T., Wiese, H., 2015. Functional gains: A crosslinguistic case study on three particles in Swedish, Norwegian and German., in: Nortier, J., Svendsen, B.A. (Eds.), *Language, Youth and Identity in the 21st Century: Linguistic Practices across Urban Spaces.* Cambridge University Press, 93–115.

Fleischman, S., Yaguello, M., 2004. Discourse Markers across Languages: Evidence from English and French, in: Moder, C.L., Martinovic-Zic, A. (Eds.), *Discourse across Languages and Cultures*. John Benjamins, Amsterdam, 129–147.

Fortescue, M., 2010. Similitude: A conceptual category. *Acta Linguistica Hafniensia 42*, 117–142.

Fox, B.A., 2010. Introduction, in: Amiridze, N., Davis, B.H., Maclagan, M. (Eds.), *Fillers, Pauses and Placeholders*. John Benjamins Publishing, Amsterdam; Philadelphia, 1–10.

Fraser, B., 2010. Pragmatic competence: The case of hedging, in: Kaltenböck, G., Mihatsch, W., Schneider, S. (Eds.), *New Approaches to Hedging*, Studies in Pragmatics. Emerald, Bingley, UK, 15–34.

Gardner-Chloros, P., 2009. *Code-switching*. Cambridge University Press, Cambridge, UK; New York.

Gardner-Chloros, P., 2013. Strasbourg revisited: *c'est chic de parler français*. *International Journal of the Sociology of Language 2013*, 143–177. https://doi.org/10.1515/ijsl-2013-0059

Graddol, D., 2010. *English Next India: The Future of English in India*. The British Council, London.

Haspelmath, M., 2010. Comparative concepts and descriptive categories in crosslinguistic studies. *Language 86*, 663–687. https://doi.org/10.1353/lan.2010.0021

Haspelmath, M., Buchholtz, O., 1998. Equative and similative constructions in the languages of Europe, in: Auwera, J. van der (Ed.), *Adverbial Constructions in the Languages of Europe*. Walter de Gruyter, 277–334.

Haspelmath, M., Tadmor, U., 2009. *Loanwords in the World's Languages: A Comparative Handbook*. De Gruyter Mouton, Berlin, Germany.

Heine, B., 2012. On polysemy copying and grammaticalization in language contact, in: Chamoreau, C., Léglise, I. (Eds.), *Dynamics of Contact-Induced Language Change*. De Gruyter Mouton, Berlin; Boston, 125–166.

Heine, B., 2018. Are there two different ways of approaching grammaticalization? in: Hancil, S., Breban, T., Lozano, J.V. (Eds.), *New Trends in Grammaticalization and Language Change*. John Benjamins Publishing Company, Amsterdam, Philadelphia, 23–54.

Heine, B., Kuteva, T., 2002. *World Lexicon of Grammaticalization*, 1st ed. Cambridge University Press.

Heine, B., Kuteva, T., 2005. *Language contact and grammatical change*, Cambridge approaches to language contact. Cambridge University Press, Cambridge; New York.

Heine, B., Kuteva, T., 2010. Contact and Grammaticalization, in: Hickey, R. (Ed.), The Handbook of Language Contact. John Wiley & Sons, Ltd, 86–105. https://doi.org/10.1002/9781444318159.ch4

Hopper, P.J., 1991. On some principles of grammaticization, in: Traugott, E.C., Heine, B. (Eds.), *Approaches to Grammaticalization: Volume I. Focus on Theoretical and Methodological Issues*, 17–35.

Hopper, P.J., Traugott, E.C., 2003. *Grammaticalization*, 2nd ed. Cambridge textbooks in linguistics. Cambridge University Press, Cambridge, UK; New York, USA.

Hurst, E., 2017. Rural/Urban Dichotomies and Youth Language, in: Ebongue, A.E., Hurst, E. (Eds.), *Sociolinguistics in African Contexts: Perspectives and Challenges*, Multilingual Education. Springer International Publishing, Cham, 209–224. https://doi.org/10.1007/978-3-319-49611-5_12

Jackson, J., 1899. *Mary Reed, missionary to the lepers*, 2nd ed. Marshall Brothers, London

Janebová, M., Martinková, M., Gast, V., forthcoming. Czech type nouns: Evidence from corpora, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move*. De Gruyter.

Jeffrey, R., Doron, A., 2013. *The Great Indian Phonebook: How the Mass Mobile Changes Business, Politics and Daily Life*. Hurst & Company: London.

Johannessen, J.B., 2012. Han professoren og sånn festskrift – om modale determinativer, in: Enger, H.-O., Faarlund, J.T., Vannebo, K.I. (Eds.), *N Grammatikk, Bruk Og Norm. Festskrift Til Svein Lie På 70-Årsdagen 15. April 2012*. Novus, Oslo, 151–172.

Jones, M.C., 2005. Some Structural and Social Correlates of Single Word Intrasentential Code-Switching in Jersey Norman French. *Journal of French Language Studies 15*, 1–23.

Kathpalia, S.S., Wee Ong, K.K., 2015. The use of code-mixing in Indian billboard advertising. *World Englishes 34*, 557–575. https://doi.org/10.1111/weng.12159

Kaltenböck, G., Mihatsch, W., Schneider, S., 2010. Introduction, in: Kaltenböck, G., Mihatsch, W., Schneider, S. (Eds.), *New Approaches to Hedging*, Studies in Pragmatics. Emerald, Bingley, UK, 1–13.

Kay, P., 1997. The kind of/sort of construction, in: *Words and the Grammar of Context*, CSLI Lecture Notes. Center for the Study of Language and Information, Stanford, California, 145–158.

Keevallik, L., 2005. The deictic *nii* 'so, in this way' in interaction., in: Monticelli, D., Pajusalu, R., Treikelder, A. (Eds.), *From Utterance to Uttering and Vice Versa: Multidisciplinary Views on Deixis.* Tartu Ülikooli Kirjastus, Tartu, 109–126.

Kilgarriff, A., Reddy, S., Pomikálek, J., Pvs, A., 2010. *A Corpus Factory for many languages.* 1-7. https://www.sketchengine.eu/wp-content/uploads/2015/05/Corpus_Factory_2010.pdf. Accessed on 31st August 2019.

Kisiel, A., forthcoming. Polish *w stylu* and the rise of hedges, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move*. De Gruyter.

Kisiel, A., Kolyaseva, A., forthcoming. Towards a comprehensive typology of type noun constructions in Slavic languages, with a special focus on Polish and Russian, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun*

Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move. De Gruyter.

Koch, P., 2016. Meaning change and semantic shifts, in: Juvonen, P., Koptjevskaja-Tamm, M. (Eds.), *The Lexical Typology of Semantic Shifts*. Walter de Gruyter GmbH & Co KG, 21–66.

Kolyaseva, A.F., 2018. The 'new' Russian quotative tipa: Pragmatic scope and functions. *Journal of Pragmatics 128*, 82–97. https://doi.org/10.1016/j.pragma.2018.03.004

Kolyaseva, A., 2021a. The nominal uses of the Russian *rod* ('genus', 'genre', 'kind') and *tip* ('type'): the starting point of desemanticization. *Slovo a slovesnost 82*, 3–44.

Kolyaseva, A., 2021b. The divergent paths of pragmaticalization: The case of the Russian particles *tipa* and *vrode*. *Journal of Pragmatics*. https://doi.org/10.1016/j.pragma.2021.08.003

Kolyaseva, A., Davidse, K., 2018. A typology of lexical and grammaticalized uses of Russian *tip*, *tipa*, *po tipu*. *Russian Linguistics 42*; Dordrecht, 191–220. http://dx.doi.org.ezproxy.lib.bbk.ac.uk/10.1007/s11185-018-9193-9

König, E., 2017. The deictic identification of similarity, in: Treis, Y., Vanhove, M. (Eds.), *Similative and Equative Constructions: A Cross-Linguistic Perspective*. John Benjamins Publishing Company, 143–164.

König, E., 2020. Beyond exophoric and endophoric uses: Additional discourse functions of demonstratives, in: Næss, Å., Margetts, A., Treis, Y. (Eds.), *Demonstratives in Discourse.* (Topics at the Grammar-Discourse Interface). Language Science Press, Berlin, 21–41.

König, E., Umbach, C., 2018. Demonstratives of manner, of quality and of degree, in: Coniglio, M., Murphy, A., Schlachter, E., Veenstra, T. (Eds.), *Atypical Demonstratives: Syntax, Semantics and Pragmatics*. Walter de Gruyter GmbH & Co KG, 285–327.

Kornfeld, L.M., forthcoming. Taxonomic nouns and markers of mitigation in Río de la Plata Spanish, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move*. De Gruyter.

Krishnaswamy, N., Burde, A.S., 1998. *The politics of Indians' English: linguistic colonialism and the expanding English empire*. Oxford University Press, Delhi.

Kuteva, T., Heine, B., Hong, B., Long, H., Narrog, H., Rhee, S., 2019. *World Lexicon of Grammaticalization*, 2nd ed. Cambridge University Press.

LaDousa, C., 2014. *Hindi is our ground, English is our sky: education, language, and social class in contemporary India*. Berghahn Books, New York.

Lehmann, C., (1982, 1995, 2002) 2015. *Thoughts on grammaticalization*, 3rd ed. Language Science Press, Berlin.

Lopes, A.C.M., Carapinha, C., 2004. Contributos para uma análise semântico-pragmática das construções com "assim." *Cadernos de Linguística 5,* 57–80.

Lopes-Damasio, L.R. UNESP, 2008. *A emergência do marcador discursivo assim sob a óptica da gramaticalização: um caso de multifuncionalidade e (inter)subjetivização.* PhD Thesis. https://repositorio.unesp.br/bitstream/handle/11449/86593/lopesdamasio_lr_me_sjrp.pdf?sequence=1

Mackerchar, E., n.d. Mary Reed of Chandag, 5th ed. *Mission to Lepers*, London.

Marques, M.A., forthcoming. Pragmatic functions and contexts of use of TIPO in European Portuguese, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move*. De Gruyter.

Marston, D., 2003. *Phoenix from the Ashes: The Indian Army in the Burma Campaign*. Greenwood Publishing Group.

Maschler, Y., 2001. *veke'ilu haraglȧyim sh'xa nitka'ot bifním kaze* ('and like your feet get stuck inside like'): Hebrew *kaze* ('like'), *ke'ilu* ('like'), and the decline of Israeli *dugri* ('direct') speech. *Discourse Studies 3*, 295–326.

Masica, C.P., 1991. *The Indo-Aryan languages*, Cambridge language surveys. Cambridge University Press, Cambridge.

Maschler, Y., 2001. veke'ilu haragláyim sh'xa nitka'ot bifním kaze (*'and* like *your feet get stuck inside* like*'*): *Hebrew* kaze *('like'),* ke'ilu *('like'), and the decline of Israeli* dugri *('direct') speech. Discourse Studies 3*, 295–326.

Matras, Y., 2009. *Language contact*, Cambridge textbooks in linguistics. Cambridge University Press, Cambridge, UK; New York.

McGregor, R.S., 2018. Introduction. In McGregor, R.S. (Ed.), 2018. *The Oxford Hindi-English Dictionary*. 58th impression. Delhi, Oxford University Press.

Mihatsch, W., 2007. The construction of vagueness: "Sort of" expressions in Romance languages, in: Radden, G., Köpcke, K.-M., Berg, T., Siemund, P. (Eds.), *Aspects of Meaning Construction*. John Benjamins Publishing, 225–245.

Mihatsch, W., 2010. The Diachrony of Rounders and Adaptors: Approximation and Unidirectional Change, in: Kaltenböck, G., Mihatsch, W., Schneider, S. (Eds.), *New Approaches to Hedging*, Studies in Pragmatics. Emerald, Bingley, UK, 93–122.

Mihatsch, W., 2016. Type-noun binominals in four Romance languages. *Language Sciences 53,* Amsterdam, 136–159.

Mihatsch, W., 2018. From ad hoc category to ad hoc categorization: The proceduralization of Argentinian Spanish tipo. *Folio Linguistica 39*, 147–176.

Mihatsch, W., 2020. A semantic-map approach to pragmatic markers: The complex approximation/mitigation/quotation/focus marking. *Marcadores Discursivos. O Português como Referência Contrastiva,* 137–162.

Mihatsch, W., forthcoming a. General introduction: Taxonomic nouns and their derived functions in Germanic, Romance and Slavic languages, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move*. De Gruyter.

Mihatsch, W., forthcoming b. A panoramic overview of the derived uses of taxonomic nouns in Romance languages, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move*. De Gruyter.

Milroy, L., Gordon, M., 2008. *Sociolinguistics: Method and Interpretation*, 2nd ed. John Wiley & Sons.

Mohan, P., 2014. The Road to English. *Economic & Political Weekly 49*, 19-24.

Montaut, A., 1995. Comparatif et comparaison dans les langues indiennes sa «quasi», de l'à-peu-près au haut degré. *Faits de langues 5*, 155–164. https://doi.org/10.3406/flang.1995.989

Montaut, A., 2015. The discourse particle *to* and word ordering in Hindi: From grammar to discourse, in: Fernandez-Vest, M.M.J., Van Valin, R.D. (Eds.), *Information Structuring of Spoken Language from a Cross-Linguistic Perspective*. De Gruyter, 263–283.

Munshi, S., 2010. Contact-induced language change in a trilingual context. *Diachronica 27*, 32–72.

Nyota, S., Mareva, R., 2012. What's New in Shona Street Lingo? Semantic Change in Lingo Adoptives from Mainstream Shona. *International Journal of English Linguistics 2*, 112–119.

Ocampo, F., 2006. Movement towards discourse is not grammaticalization: The evolution of *claro* from adjective to discourse particle in spoken Spanish, in*: Selected Proceedings of the 9th Hispanic Linguistics Symposium*. Cascadilla Proceedings Project, Somerville, MA, 308–19.

Odden, O.R., 2019. *North Scandinavian type noun constructions: Patterns with slags, SORTs and TYP(E).* Unpublished PhD thesis. University of Oslo.

Oesch Serra, C., 1998. Discourse Connectives in Bilingual Conversation: The case of an emerging Italian-French mixed code, in: Auer, P. (Ed.), *Code-Switching in Conversation: Language, Interaction and Identity*. Routledge, London; New York, 101–121.

Onysko, A., Winter-Froemel, E., 2011. Necessary loans-luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics 43*, 1550–1567.

Overstreet, M., 2005. And stuff und so: Investigating pragmatic expressions in English and German. Journal of Pragmatics, *Approaches to Spoken Interaction 37*, 1845–1864. https://doi.org/10.1016/j.pragma.2005.02.015

Overstreet, M., 2014. The role of pragmatic function in the grammaticalization of English general extenders. *Pragmatics 24*, 105–129.

Parshad, R.D., Bhowmick, S., Chand, V., Kumari, N., Sinha, N., 2016. What is India speaking? Exploring the "Hinglish" invasion. *Physica A: Statistical Mechanics and its Applications 449*, 375–389. https://doi.org/10.1016/j.physa.2016.01.015

Peterson, E., 2017. The nativization of pragmatic borrowings in remote language contact situations. *Journal of Pragmatics 113*; Amsterdam, 116–126.

Pichler, H., Levey, S., 2011. In search of grammaticalization in synchronic dialect data: general extenders in northeast England. *English Language & Linguistics 15*, 441–471. https://doi.org/10.1017/S1360674311000128

Portner, P.H., 2005. *What is Meaning? Fundamentals of Formal Semantics*. Wiley.

Rosenkvist, H., Skärlund, S., 2013. Grammaticalization in the present - The changes of modern Swedish typ, in: Ramat, A.G., Mauri, C., Molinelli, P. (Eds.), *Synchrony and Diachrony: A Dynamic Interface*. John Benjamins Publishing, 313–340.

Rosignoli, A., 2011. *Flagging in English-Italian code-switching*. Unpublished PhD thesis. Prifysgol Bangor University.

Sahoo, K., van der Auwera, J., 2019. South-Asian similatives: A typological perspective. *Lingua 226*; Amsterdam, 20–34. http://dx.doi.org/10.1016/j.lingua.2019.05.006

Salkie, R., 2002. Two types of translation equivalence, in: Altenberg, B., Granger, S. (Eds.), *Lexis in Contrast: Corpus-Based Approaches*. John Benjamins Publishing, 51–72.

Sayers, D., 2014. The mediated innovation model: A framework for researching media influence in language change. *Journal of Sociolinguistics 18 (2)*, 185–212. https://doi.org/10.1111/josl.12069

Schiering, R., 2010. Reconsidering erosion in grammaticalization, in: Stathi, K., Gehweiler, E., König, E. (Eds.), *Grammaticalization: Current Views and Issues*. John Benjamins Publishing, Amsterdam; Philadelphia, 73–100.

Schoning, C., Helder, J., Diskin-Holdaway, C., 2021. *Vi snakker sådan*: An analysis of the Danish discourse-pragmatic marker *sådan*. *Nordic Journal of Linguistics*, 1–37. https://doi.org/10.1017/S0332586521000159

Si, A., 2011. A diachronic investigation of Hindi-English code-switching, using Bollywood film scripts. *International Journal of Bilingualism 15*, 388–407.

Stanlaw, J., 2004. *Japanese English: language and culture contact*. Hong Kong University Press, Aberdeen; Hong Kong.

Terraschke, A., Holmes, J., 2007. 'Und Tralala': Vagueness and General Extenders in German and New Zealand English, in: Cutting, J. (Ed.), *Vague Language Explored*. Palgrave Macmillan UK, London, 198–220. https://doi.org/10.1057/9780230627420_11

Thomason, S.G., 2001. *Language contact: an introduction*. Edinburgh University Press, Edinburgh.

Thompson, C.P., 2019. *Tipo*: Innovation and change in Brazilian Portuguese. *Diacrítica 33*, 196–215. https://doi.org/10.21814/diacritica.421

Traugott, E.C., 1995. The role of the development of discourse markers in a theory of grammaticalization. *Unpublished paper presented at the 12th International Conference on Historical Linguistics*, Manchester.

Traugott, E.C., Dasher, R.B., 2001. *Regularity in Semantic Change*. Cambridge University Press.

Treis, Y., Vanhove, M., 2017. *Similative and Equative Constructions: A cross-linguistic perspective*. John Benjamins Publishing Company.

Trudgill, P., 1988. Norwich Revisited: Recent Linguistic Changes in an English Urban Dialect. *English World-Wide 9*, 33–49. https://doi.org/10.1075/eww.9.1.03tru

Umbach, C., forthcoming. Means of classification: German *Art* and *Typ*, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move*. De Gruyter.

Umbach, C., Gust, H., 2014. Similarity demonstratives. *Lingua 149*, 74–93. https://doi.org/10.1016/j.lingua.2014.05.003

Umbach, C., Gust, H., 2021. Grading similarity, in: Löbner, S., Gamerschlag, T., Kalenscher, T., Schrenk, M., Zeevat, H. (Eds.), *Concepts, Frames and Cascades in Semantics, Cognition and Ontology*. Springer, Cham, Switzerland, 365–386.

Vaish, V., 2008. Language Attitudes of Urban Disadvantaged Female Students in India: An Ethnographic Approach. *Journal of Multilingual & Multicultural Development 29*, 198–215.

van der Auwera, J., Coussé, E., 2016. Such and *saadan*–the same but different. *Nordic Journal of English Studies 15*, 15–32.

van der Auwera, J., Sahoo, K., 2020. Such similatives: a cross-linguistic reconnaissance. *Language Sciences 81*, 1–13. https://doi.org/10.1016/j.langsci.2018.12.002

Vassiliadou, H., Vladimirska, E., Lammert, M., Benninger, C., Gerhard–Krait, F., Gridina, J., Turla, D., forthcoming. Clear vs approximate categorization in French and Latvian, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), *Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move*. De Gruyter.

Vindenes, U., 2017. Complex demonstratives and cyclic change in Norwegian. Unpublished PhD thesis. University of Oslo, Oslo.

Voghera, M., 2013. A case study on the relationship between grammatical change and synchronic variation, in: Ramat, A.G., Mauri, C., Molinelli, P. (Eds.), *Synchrony and Diachrony: A Dynamic Interface*. John Benjamins Publishing, 283–312.

Voghera, M., forthcoming. The network of specie, genere, sorta, tipo constructions: from lexical features to discursive functions, in: Brems, L., Davidse, K., Hennecke, I., Kolyaseva, A., Kisiel, A., Mihatsch, W. (Eds.), Type Noun Constructions in Slavic, Germanic and Romance Languages: Semantics and Pragmatics on the Move. De Gruyter.

Wiese, H., 2011. So as a focus marker in German. *Linguistics 49*, 991–1039. https://doi.org/10.1515/ling.2011.028

Zalizniak, A.A., 2018. The Catalogue of Semantic Shifts: 20 Years Later. *Russian Journal of Linguistics 22*, 770–787.

Dictionaries

Kumar, S., Sahai, R., 2014. *Oxford English-English-Hindi Dictionary*. 2nd Edition. Delhi: Oxford University Press.

McGregor, R.S., 2018. (ed.) *The Oxford Hindi-English Dictionary*. 58th impression. Delhi: Oxford University Press.

Websites

Lewis, M. Paul, Gary F. Simons, and Charles D. Fennig (eds.). 2016. *Ethnologue: Languages of the World*, Nineteenth edition. Dallas, Texas: SIL International. Available at http://www.ethnologue.com/profile/IN. Last accessed 18 May, 2016.

Office of the Registrar General & Census Commissioner, India: Ministry of Home Affairs, Government of India http://www.censusindia.gov.in/2011-prov-results/paper2/prov_results_paper2_uttrakhand.html. Last accessed 18 June, 2021.

Max Planck Institute for Evolutionary Anthropology, Department of Linguistics. *Leipzig Glossing Rules.* https://www.eva.mpg.de/lingua/resources/glossing-rules.php Last accessed 12 May 2017.

Historical Thesaurus of the Oxford English Dictionary (HTOED) https://www.oed.com/view/Entry/208330? Accessed 22 January, 2020.