

Tatiana Oranskaia 

Hindi Clause Strings with Adverbial Clauses: A Tentative Formalisation

तत्याना ओरांस्कया

हिंदी की उपवाक्य शृंखलाएँ – अंतरिम सूत्रबद्ध प्रणाली

Abstract The article presents a tentative formalisation system for multicomponent Hindi clause strings which include at least one adverbial clause. The aim of the study is to create a tagged database of such constructions. The relations between the members of “classical” biclausal constructions, subordinate and coordinate, serve as the benchmarks for investigating longer clause sequences occupying a level intermediary between complex, compound and compound-complex sentences on the one hand and larger text units on the other hand. The special interest in the subordinate modifier clauses is caused by their semantic variety and relatively loose—as compared to complement clauses—connection with the main clause. Due to the latter characteristic adverbial clauses offer more points of syntactic similarity with simple sentences, and so a better possibility for testing syntactic relations between members of a clause string as well as their semantic grounds. Only finite clauses are considered here components of clause strings. The formalisation is expected to on the one hand explicate the results of the preliminary analysis and on the other hand facilitate further exploration on the topic.

Keywords subordination, coordination, formalisation system, gradience, structural variance.

सारांश प्रस्तुत लेख हिंदी की ऐसी बहुअवयवी उपवाक्य शृंखलाओं (multicomponent clause strings) के अध्ययन का प्रयास है जिनमें कम-से-कम एक क्रियाविशेषण उपवाक्य विद्यमान हो। इस तरह के उपवाक्यों की एक अंतरिम सूत्रबद्ध प्रणाली (tentative formalisation system) यहाँ प्रस्तुत की गई है। हमारा उद्देश्य इस तरह की वाक्य-रचनाओं का चिह्नित डाटाबेस तैयार करना है। प्रधान और आश्रित उपवाक्यों से बनी 'पारंपरिक' (classical) वाक्य-रचनाओं के अंगों (members) के पारस्परिक संबंधों को आधार बनाकर ऐसी उपवाक्य शृंखलाओं का अलग से अध्ययन किया जा सकता है जो, एक ओर तो मिश्र, संयुक्त तथा संयुक्त-मिश्र वाक्यों की दृष्टि से तथा दूसरी ओर संपूर्ण पाठ जैसी अधिक बड़ी इकाई की दृष्टि से „मध्यवर्ती” (intermediary) संरचनाएँ मानी जा सकती हैं। आश्रित विशेषक उपवाक्य (subordinate modifier clauses) हमारा ध्यान इसलिये आकृष्ट करते हैं क्योंकि उनमें अर्थगत विविधता होती है तथा प्रधान उपवाक्य के साथ, पूरक उपवाक्यों की तुलना में, उनका संबंध अपेक्षाकृत शिथिल (loose) होता है। प्रधान उपवाक्य के साथ इस अपेक्षाकृत शिथिल संबंध के कारण क्रियाविशेषण उपवाक्यों की वाक्यरचनागत समानता सरल वाक्यों के साथ अधिक दिखाई देती है और इसलिये ऐसे उपवाक्यों के आधार पर उपवाक्य शृंखलाओं के अंगों के बीच विद्यमान वाक्यगत संबंधों और उनकी विशेषताओं की जाँच करना संभव प्रतीत होता है। यहाँ पर केवल समापिका क्रियायुक्त उपवाक्यों (finite clauses) को ही उपवाक्य शृंखलाओं का अवयव (component) माना गया है। उपवाक्य शृंखलाओं को सूत्रबद्ध करने (formalisation) के इस प्रयास के द्वारा प्रारंभिक विश्लेषण के परिणाम कुछ अधिक स्पष्ट होने के साथ-साथ इस अध्ययन को आगे बढ़ाने में भी सहायता मिल सकेगी।

मुख्य शब्द – आश्रितता/अधीनता, समानाधिकरण, सूत्रबद्धीकरण प्रणाली, ग्रेडिएन्स (gradience), संरचनात्मक अंतर।

1 Introduction

The article presents a tentative version of a system to formalise clause strings in Modern Standard Hindi and the analysis underlying the formalisation. It is the first published outcome of my research project “Hypotaxis in Spoken and Literary Hindi: A Comparative Analysis of Complex Sentences with Adverbial Clauses” at the Mahatma Gandhi Antarrashtriya Hindi Vishwavidyalaya (Mahatma Gandhi International Hindi University) in Wardha, Maharashtra.¹ Its eventual objective is to create a tagged database for complex syntax in Hindi focusing on adverbial clauses.

- 1 I am deeply grateful to the faculty and staff of this—in spite of its young age—renowned university for the wonderful opportunity to work there and benefit from kind assistance of a number of colleagues as well as from the vibrant atmosphere of the university. I am greatly obliged to the Indian Council of Cultural Relations whose fellowship under the programme “Fellowship to Foreign Nationals for Research in India” enabled my research stays in India in 2016–2018. My heartfelt thanks are due to ex-Vice Chancellor Prof. Girishwar Misra, ex-Deputy Vice Chancellor Prof. Anand Vardhan Sharma, Dr Usha Sharma, Dean of the School of Language Prof. Hanuman Prasad Shukla and Prof. Anil Kumar Pandey for their kind attention to my research efforts and for facilitat-

The article is split into two parts. The first part is subordinate, the second part is the main one, to use two of the key concepts in the present discussion. The first part contains an introductory review of issues relevant to the suggested formalisation system. The main part presents a tentative formalisation of clause strings and the underlying analysis. It is a linear tectogrammatical presentation of interclausal relations, refraining from an analysis of the interior syntax of the clauses. No hierarchical trees are built.

This is a result of the first stage of the project, in which the data are elicited from written sources. The next stage of the study will be based on spoken language data.

The definition “clause strings” refers to clause sequences consisting of more than two clauses. Such long strings occupy a level intermediary between complex and compound sentences on the one hand and larger text units on the other hand. Interpretation of these structures coincides with the problem of information hierarchy in human language, which includes the issue of subordinating devices.

We can imagine a long clausal string as a sequence of links, each consisting of two clauses and a binding link carrying the meaning of the semantic connection between them. The dependency distance between the constituent clauses is a variable related to the semantic type of their interconnection. The binding link is not necessarily formally explicit: the connection may rely on implicit meanings binding the clauses together.

“We construe the same situation in alternative ways”, to cite Langacker (2010: 55). The more complex the situation is, the more propositions participate in portraying it. Accordingly, the number of clauses framing them is higher, which allows a higher variation in their linkage leading, in its turn, to a variation increase in perspectives on the situation.

Multicomponent-sequences, which include at least one adverbial clause, are the focus of the project. The adverbial interclausal relations belong principally to the domain of subordination.

As an umbrella notion, complex syntax shares part of its domain with the syntactic organisation of larger text chunks where sentences may have no other form

ing them. Very special thanks are due to Dr Shamim Fatma and Dr Dhanjee Prasad from whose scholarly knowledge, skills and active interest in my topic I have immensely benefited. I cannot thank enough the bright and enthusiastic students who participated in the project. It is with a heavy heart that, due to lack of space, I have to leave them anonymous and also exclude some other parts of my gratitude list, such as my heartfelt thanks to the staff of the Guest House. It is my pleasant duty to thank Dr Joan M. Barry and Mr Philip H. Pierce profoundly for their insightful corrections of my English and Dr Hem Chandra Pande for patiently answering my innumerable questions on linguistic terminology in Hindi and other issues.

to express their semantic connections but the adjoining position—and sometimes not even that.

With the best will in the world complex syntax cannot be considered a new research subject. However, the discussion on clause complexing remains profuse and intense, gaining new force since the 1960es. The subject has been studied in different theoretical frameworks using various data types and formats (see, e.g. Haiman & Thompson 1988; Shopen 2007). As a resource for constructing our experience of the “flow of events”, it is especially significant for the understanding of human cognitive activities, discourse structuring and its hierarchy. Especially promising in this respect are studies on spoken language syntax based on substantial data collections made possible by modern technical devices (Auer et al. 2009).

Adverbial clauses have been attracting considerable attention from syntacticians (see Thompson, Longacre & Hwang 2007). One of the reasons is their relatively loose link to the matrix clause, which is a significant fact for the debate on gradience in grammar (Fanselow et al. 2006). The cross-linguistic semantic and syntactic heterogeneity of adverbial modifiers in general (see, e.g. Eifring 1995: 54; Ricca 2010) is the reason why adverbial clauses span over the subordination–coordination axis.

Against this rich and multifaceted background, research on complex syntax in Hindi is rather scarce. Studies on subordination in Hindi deal for the most part with complement and noun-modifying relative clauses (e.g. Ananthanarayana 1996; Bhatt 2003; Kachru 1978; Dayal 1996; Kothari 2010). Adverbial interclausal relations are to a considerable extent neglected. Accordingly Hindi data are only occasionally used in typological research and on the periphery of scholarly discussion. Thus, Hindi is not considered in the generalising works on syntax mentioned above (Haiman & Thompson 1988; Shopen 2007), nor is it among the 60 languages involved in the crosslinguistic study on causal clauses (Diessel & Hetterle 2011).² Some attention has been given to conditional constructions in Hindi (Oranskaya 2005; Sharma 2010; 2012).

Naturally, analyses of complex syntax underlie implicitly or explicitly annotations in Hindi databanks, such as for example in the Hindi Discourse Relation Bank (Umangi et al. 2009) or the annotation in The Hindi/Urdu Treebank Project (Bhat et al. 2017). Another example is the annotated corpus data on relative clauses contained in the Appendix A to the PhD thesis of Anubha Kothari (Kothari 2010).

This article is a step towards a tentative version of a formalisation system for complex syntax in Hindi. It concentrates on interclausal relations with the adverbial semantics. The formalisation system is expected to enable a closer look into

2 Of Indian languages only Santali and Kannada are considered in this study.

variations in the ways to connect clauses, the placement of adverbial clauses in clause sequences, and the scope of the latter.

2 Basic concepts and terms

The analysis and, accordingly, the formalisations of clause strings in 5.2 use traditional concepts. The major concept is clause: it is a structure including at least a subject and a predicate. Predicate is understood here as a finite predication phrase. This is different from the usual approach, which relates also infinite verb forms to clausal predicates (cf., e.g. Lehmann 1985). Homogeneous subjects and homogeneous predicates are considered to belong to the same clause.

Clause is a relative notion determined by syntactic context. A structure of any syntactic complexity is a clause if it is itself a part of a composite unit whose integrity is based on semantic and syntactic relations within it. This interpretation of the term follows the definition of sentence as a combination of clauses (Longacre 1970; 2007). A sentence may also contain just one clause.

Further concepts belong to the sphere of complex syntax. They are listed here according to their complexity, starting with the simplest one.

C o m p o u n d s e n t e n c e is a composite syntactic unit whose constituent clauses are in a structural equivalence relation, that is **c o o r d i n a t i o n**.

C o m p l e x s e n t e n c e is a composite syntactic unit whose constituent clauses are in a structurally hierarchical relation, that is **s u b o r d i n a t i o n**.

C o m b i n e d s e n t e n c e: the term is reserved here for clause strings that conjoin more than two clauses using explicit devices of both subordinating and coordinating types. The phrase “explicit devices” refers almost exclusively to formal lexico-grammatical markers and to a few semantically based common types of clause binding, for example, attitude verbs. Otherwise, a semantically motivated clause string with an unmarked clause adjunction does not qualify as a sentence.

C l a u s e s t r i n g may be a sentence of one of the types characterised above or a sequence of clauses whose semantic interrelations do not necessarily receive an expression through lexico-grammatical means.

The notion of clause string raises the question of its right boundary. The question of boundaries is typically a tricky one. Most difficult to overcome in research on the syntax of spoken language (Auer 1992: 41), it also presents enough difficulties in syntactic exploration of written texts.

In functional linguistics, a sentence of a written text is defined on the basis of graphological features. As a rule, in texts written in such scripts as Roman or Cyrillic, the beginning of a sentence is marked by an uppercase letter and punctuation marks are generally conceived as signs corresponding to the prosodic

signs of intonation units in oral communication (Chafe 1984). However, the Devanagari script (used by Hindi) does not distinguish between uppercase and lowercase letters. Punctuation in Hindi is sparser than in texts written in the most widely-used letter scripts. Although all punctuation marks of these scripts are also used in Devanagari the sign for a full stop—a short vertical line (।)—often also occurs in syntactic positions, where a European text has a comma or another sign for marking a syntactic unit. Just as such punctuation usage complicates defining the sentence boundaries, so it may produce a yet stronger variation in interclausal parsing. If a clausal string does not fit into any definition of a sentence, this study draws on the semantic interrelations characterising a multiple event sequence as the ground for delimitation of strings. The criterion is, of course, anything but accurate. However, on the whole it seems to work. Propositional relations build a by and large usable foundation for clause-by-clause parsing.

Coming to *c o o r d i n a t i o n* and *s u b o r d i n a t i o n*, the broadest concepts relevant for the study, we find that their opposition is somewhat problematic. Not only do constructions show mixtures of subordination and coordination (Haspelmath 2004: 37), but both types of clause-linkage may overlap in expressing the same meaning. Owing to the semantic and formal multifariousness of interclausal bonds, the status of subordination as a grammatical category in its own right is placed in doubt (Cristofaro 2014; Herlin et al. 2014).

A more persuasive approach views subordination as gradient along the opposition axis whose other pole is coordination. It is an established idea that grammatical categories are gradient (Fanselow et al. 2006; Traugott & Trousdale 2010) and there is no reason why subordination would differ from other syntactic phenomena. For the weaker grade of the hierarchical dependence in a sentence, i. e. not embedded subordination, Foley & Van Valin (1984, ch.6) use the term “cosubordination”.

The array of phrasal and adpositional modifiers to the terms “coordination” and “subordination” speaks clearly of their terminological insufficiency. Compare, for example, “syndetic” and “asyndetic” applied to both the terms (Jucker 1991), “pseudo-coordination” (Ross 2016), “genuine coordination” (Ledgeway 2016: 157) or the term “insubordination”, which appeared at the break of the 21st century and immediately gained strength in linguistic theory (Heine et al. 2016).

For all its imprecision, subordination is an unavoidable concept, convenient as a contrast to “coordination” (among other things). There is no other term that can be deployed when discussing the hierarchical organisation of clausal units within clause strings.

3 Adverbial clauses

Adverbial clauses share specific characteristics with lexical adverbials. Most important is that semantic factors dominate over grammatical in determining the position of adverbial clauses in complex syntactic hierarchies. It has been demonstrated for English that the semantic content of interclausal relations plays the major role in the ordering of main and adverbial clauses (Diessel 2005).

The grade of variation is in direct proportion to the scope of grammatical freedom. The syntactic heterogeneity of clause-linkages with adverbial semantics goes along with a relatively loose connection between adverbial and main clauses (Chafe 1984), or “loose subordination” (Givón 1990). Adverbial clauses are characterised by a “low degree of integration into the matrix clause... and a low degree of interlacing”, as it is the case in the core languages of Europe (Kortmann 1997: 241). We can add also Hindi, insofar as it concerns conjoined clauses with finite verb forms.

The dependence distance between an adverbial and the superordinate clause is variable. Another specificity is that an adverbial clause’s governor may be a complex multiclausal structure of which it itself is a part. Semantic variety of adverbial clauses is combined with diverse syntactic marking. In other words, adverbial clauses show a gradience in the degree of subordination. This makes the task of presenting their relations within long clause strings through a formalised coding system look like a promising method of capturing their basic syntactic features and idiosyncrasies.

Various perspectives on characteristics of adverbials have been summarised by Ernst (2020). Among the subjects of discussion are their location, distribution, correlation between semantic and syntactic factors, etc. Formalisation, especially in the form of a database, makes it easier to capture such issues and the variations in framing various communicative strategies.

4 Data and principles of analysis

The formalisation presented in this article is based on data from written sources. The major part of the processed data stems from essays of modern Hindi writers accessible on the Internet (<http://hindisamay.com>).³ The choice of the genre of essay is due to its closer connection to the reality lived by the author. Its fragmented, predominantly monologic form and incoherent composition show an unconventionally strong personal influence of the author on the form and content

3 As the texts placed on the website had been typed out from books and magazines, diversions from the originals are possible.

(Wang & Jan 2018: 296). Because of these and some other, less relevant, features essay is closer to the spoken language than other literary genres. This is especially strongly felt in the syntactic characteristics of these texts. In them, utterances recorded in the written form are characterised by a comparatively free form of assemblage of clauses and a strong tendency to build long clause sequences, thus encompassing on the average a higher number of mutually related micro-situations than a sentence in a text of a higher literary level. Syntactic and semantic connections within a string can spread over distant clauses, resulting in a portrayal of a multidimensional complex situation beyond the scope of complex and compound sentences.

As of this stage of the study, all respondents have been educated Hindi native speakers from various dialect backgrounds. Students of the School of Language, Mahatma Gandhi Antarrashtriya Hindi VishwaVidyalaya, were assisting in data collection and in preparing a database. Further respondents have given occasional assistance in the former task. Many of the students participated in a six-day training workshop *Hindī miśr vākyō mē ṭaigḍ ḍeṭābes nirmāṇ* (Creation of Tagged Database for Hindi Complex Sentences; 12–17 February 2018). Working together with the students opened to me new vistas and resulted in a number of changes introduced since then in the formalisation system. At this point I heartily thank the students once again. The opportunity to enjoy this collaboration I mostly owe to two of their teachers, Dr Shamim Fatma and Dr Dhanjee Prasad. It was they who came up with the idea of organising such a workshop and did a superb job of bringing it to life. Needless to say, their contribution to the project was not limited to the organisational aspect.

The obtained data bring to light strong structural variations in expressing the same logico-semantic structures. Hypotactic and paratactic clause complexes alternate in their nexus meanings with each other and with sequences of simple sentences. Such alternations can hardly be free considering that the broadest scope of logico-semantic relations obtains at the level of clause complexes (Halliday 1985).

Tectogrammatical representations of clause strings have been developed with a view to creating an interactive database. For representations see 5.2 below.

At the initial stage the formalisation system is based on an analysis of a limited amount of data and is being developed by involving larger data. At the same time, its development serves to elaborate and correct the methods and results of the analysis.

The analysis underlying the formalisation proceeds from biclausal to multi-clausal sequences. The benchmark in the analysis is clausal units which incorporate two clauses and a semantic link between them. Of their two basic types—compound sentence and complex sentence—the latter, comprising a main clause and a subordinate adverbial clause, is more significant for the discussion here.

However, the adverbial semantics of the link between two clauses can in a number of cases be expressed by compounding them, exposing a partial synonymy of hypotactic and paratactic constructions. Moreover, the semantic relation can also exist between juxtaposed sentences which are otherwise syntactically independent (Aguiar & Barbosa 2016: 12).

The variety of expression of interclausal adverbial relations reveals their strongly gradient character.

5 Formalisation system

A string is a hierarchical structure. Nevertheless, formalisations are structured horizontally in order to reflect the unfolding of strings along the time axis in spoken and written language forms and in accordance with graphic presentation of a language in a left-to-right script, which is the case with Hindi written in Devanagari.

The preliminary variant of the formalisation presented here is being developed and expanded with new data. Search for an optimal formalisation facilitates the analysis. The procedure accepts the standpoint that strings are conveniently analysed as a linear structure (Longacre 1960).

The tagging procedure follows a major principle of Natural Language Processing, according to which annotations should not alter the underlying corpus in any way; that is, tags are separated from the data using them. This is known as the principle of *stand-off annotation* (Ide & Romary 2004).

Four steps precede the tagging procedure:

- (a) Parsing into clauses
- (b) Disambiguation of interclausal meanings
- (c) Disambiguation of intersentential meanings
- (d) Establishing types of clause combining.

In the schematic presentation all subordinating connectors are placed in the subordinate clauses and coordinating connectors are placed in the matrix clause. This principle is also deployed in the clause-by-clause parsing of the strings. This strategy is tentative, adopted for mere convenience. Clausal affiliation of various connecting devices needs to be explored.

A range of clause binding means corresponds to the semantic heterogeneity of adverbials, from marked subordination through compounding to juxtaposing clauses which are formally independent sentences.

In order to capture the formal variety of interclausal relations five types of brackets are used along with other tags. It is a peculiarity of the system. Other linear presentations of syntactic constructions use two or at most three types: round,

square and angled brackets. So, Langacker (2014) uses three types of brackets and additionally slashes and double slashes alternately for presenting clausal and phrasal structure of sentences linearly and the greater-than sign for establishing interclausal hierarchy in asyndeton. In linear tectogrammatical schemes he combines round and square brackets. Three types are also used in the Transcription System of Spoken Language (Auer et al. 2009). The annotations in the Hindi Discourse Relation Bank use square and curly brackets to mark the ordering of clauses (Umangi et al. 2009). Kothari (2010) uses round brackets to demarcate clauses and square brackets for morphosyntactic tags. Tree-form annotation is superfluous to this review.

The tools used in the tectogrammatical formalisation described below are presented in Table 1. Most syntactic tags are common for linear syntactic annotations. Some tags occur here for the first time. In any case, I never came across them in the literature on the topic. Some semantic tags used here are mine, four are borrowed from English Propbank (Bonial et al. 2015).

5.1 Formalisation tools and tagging guidelines

Table 1 Brackets

| |
|--|
| <p>Angle brackets < ></p> <p>Tags for each clause (also when a clause is a sentence) are given in angle brackets.</p> |
| <p>Round brackets ()</p> <p>Tags for all composite (non-simple) sentences are taken in round brackets when a composite sentence occurs in the string for the first time (unlike U-brackets, see below). A left round bracket introduces each compound, complex and combined sentence. The number of round brackets on the right boundary of a tectogrammatical scheme corresponds to the number of conjoined sentences.</p> |
| <p>Square brackets []</p> <p>Sequences of clauses forming a compound or complex sentence are given in square brackets.</p> |
| <p>Curly brackets { }</p> <p>All tags with associated lexical semantics, including terminology, are given in curly brackets. These are conjunctions and all functional expressions used to combine clauses, or else are denotations of semantic relations between clauses. The conjunctions and other lexical items with syntactic functions are italicised in the schemes.</p> |
| <p>U-brackets ⊂ ⊃</p> <p>Tags for composite sentences and connectors participating in more than one structural relation are given in U-brackets, when not manifest in the surface structure. Repeatedly used sentences are indicated by the same subscript letter.</p> |

Table 2 Terms, tags and tagging examples

| | |
|---|---|
| <p>Terms and their tags</p> | <p>Tagging examples</p> <p>The examples here are from the essay of Ajayendranāth Trivedī <i>Bar̥kā jāmun</i> <www.hindisamay.com/contentDetail.aspx?id=5514&pageno=1>. In other cases reference is given to the formalisations in 5.2.</p> |
| <p>Clause <C> Main clause <MC> Subordinate clause <SC> Complex sentence <CxS></p> | <p>(<CxS {tmp}>[<i>tez havā caltī</i> <SC> Strong wind blew <SC> <i>to kathjāmun jamīn par bich jāte</i> <{to} MC>]) then java.plums ground on spread went<{then}MC>]) '[When] strong wind was blowing the java plums were raining to the ground.'</p> |
| <p>Compound Sentence Clauses in a source text passage are numerically indexed according to their sequence in the source text passage: C1, C2 ... Cn</p> | <p><CpS> (<CpS {aur} {tmp}> [subah hotī <C1> {aur} [morning came <C1> {and} <i>usī bar̥kā jāmun ke tale dhān kī</i> that.very big java.plum GEN under rice GEN <i>piñī śurū hotī</i> <C2>]) threshing beginning was <C2>]) 'Morning came and under the same big java plum tree began threshing of rice.'</p> |
| <p>Combined sentence</p> | <p><CdS> See for an example 5.2 (D).</p> |
| <p>Semantics of the relations between clauses; the conjunctions are italicised in the schemes.</p> | <p>temporal {tmp}, location {loc}, cause {cau}, effect {eff}, {cau-eff}, conditional {cnd}, resultative {res}, purpose {prp}, concessive {cnces}, consecutive {cnsect}, complement {cmpl}, quotation {quot}, consequence {cnseq}, manner {mnr}, restriction {rstr}, comparison {cmpr}, attributive {attr}</p> <p>Meaning concretisation is expressed by a colon (:) before the tag extension, e.g. {tmp: immediate sequence}</p> |
| <p>When conjunctions are shifted from their initial position to a position inside the clause, they are marked in the scheme by hyphens on both sides of the tag, (as -{yadi}- in the example here).</p> | <p>(<CxS {cnd-res} {yadi-to} [<SC {cnd} -{yadi} -> <i>apnā patā-ṭhikānā yadi kisī=ko</i> own address-living.place if somebody=DAT <i>batānā ho</i> <{to}MC {res}>]) <i>to bas</i> tell.INF be.CONJ.3SG then just <i>itnā hī kahnā hamāre lie kāfī thā</i> ... that.much only say.INF us for enough was ... 'If we had to explain somebody where we live, it was enough to say just...'</p> |
| <p>Conjunctions in their usual position at the beginning of a clause</p> | <p><{to} MC>, <{tab} MC> <{agar} SC>, <{jab} SC></p> |
| <p>Clause order</p> | <p>[<SC> < MC>] or [<MC> <SC>] or [<MC -<SC>->] The latter scheme presents embedded SC.</p> |

| | |
|--|---|
| Terms and their tags | <p>Tagging examples</p> <p>The examples here are from the essay of Ajayendranāth Trivedī <i>Barkā jāmun</i> www.hindisamay.com/contentDetail.aspx?id=5514&pageno=1. In other cases reference is given to the formalisations in 5.2.</p> |
| <p>Clause valence – the term denotes the number of clauses with which a clause is syntactically connected. If a clause is involved in syntactic relations with more than one clause, it is assigned a valence corresponding to the number of clauses connecting to it. The number of the valence tag is the sequence number of the valence realisation in the string linear structure. The valence tag is parenthesised together with its clause tag and separated from it by a colon.</p> | <p>val1, val2, val3 [<MC:val1><{ki}SC>] See for an example 5.2 (A).</p> |
| <p>Level – the letter “L” with a following number denotes the level in the hierarchical structure of the strings.</p> | <p>[{<jab/cause}SC-L3><{to}MC-L2:val1>] See for an example 5.2 (C).</p> |
| <p>An elucidating sentence depicting a situation reflected in the analysed clause(s) is given in double slashes //.</p> | <p>//Hamāre barkā jāmun ko kaun nahī jāntā// //‘Who doesn’t know our big java plum tree?’//</p> |

- Clause tags are positioned after each clause, i. e. after its number and before punctuation signs.
- If an embedded clause is placed inside the matrix clause, the initial part bears the clause number with the postposed number sign (#), whereas the bare number appears at the end of the clause.
- In order to distinguish main clauses and subordinate clauses of different levels in a string hierarchy, each clause is indexed with its level number, e.g. MC-L2 means “main clause of the 2nd level”.
- The subscript numbers in the data refer to the clauses. Along with the numbering, they mark the right boundaries of the clauses. In the text of the article the subscript numbers are substituted with bracketed numbers.⁴
- Tectogrammatical structures are positioned after the clause strings.

4 In the continuation of the studies the clause count will be used in a quantitative analysis.

5.2 An analysis and a tentative formalisation of clause strings

This part includes an analysis of strings in a passage from an essay by Buddhināth Miśra (Miśra s. a.) *Phūl āe haī kanerō mē* ('Oleanders are Blossoming'). I selected this passage because it includes four clause strings which build an almost uninterrupted sequence and thus present a convenient opportunity to explore the transitional level between sentence syntax and text syntax. As stated above, the style characteristics of the essay genre are, as a rule, fairly close to those of oral narrative, which is the case in the analysed extract. In this part of the text, the key figure is not the author, an uncommon characteristic of essays. This is about a person who found himself in an unknown village and asked for shelter for the night in a house which, like his own house, had a jujube tree in front of it. The host on his return home from the fields found an unknown person there who had introduced himself to the family as their relative. Next morning the host asked the guest about his place of residence and their relationship. Clause strings (A)–(C) are parts of the conversation corresponding to its timeline. Clause string (D) precedes the conversation in narrative time. However, it is placed last in the analysis in order to separate it from the strings which are considered sentences.

- (A) *ham jānnā cāhte haī₁ ki āp kis gāv ke sambandhī haī₂,*
 We know want AUX₁ that you what village of relation are₂
kyōki āj tak hamne kabhī āpko dekhā nahī₃.
 because today until we somewhen you.OBJ saw not₃.

($\langle \text{CxS}\{\text{cmpl}\}\{\text{cause}\} \rangle \langle \text{CxS}_i\{\text{cmpl}\} \rangle [\langle \text{MC}_1:\text{val1} \rangle \langle \{ki\}\text{SC}_2 \rangle] \langle \text{CxS}\{\text{cause}\} \rangle$
 $[\langle \langle \text{CxS}_i [\langle \text{MC}_1:\text{val2} \rangle \langle \{ki\}\text{SC}_2 \rangle \rangle \rangle] \langle \{kyōki\}\text{SC}_3 \rangle]]]$)

'We want to know₁, from what village are you our relation₂,
 because until today we had never seen you₃.'

String (A) is an exemplary case of an adverbial connection between two syntactic units, the first of which is a biclausal complex sentence and the second a clause carrying the adverbial meaning. In traditional terminology it is a complex sentence with two subordinate clauses, whereas the deeper structural relations need further comments. Both subordinate clauses depend on the same main clause. The dependencies within the string are asymmetrical. Clause (2), a complement of the verb *jānnā* 'to know', is embedded in clause (1), the connection being marked by the complementiser *ki* (from Persian, lit. 'who', 'which', 'why'), approximately corresponding to 'that' but, unlike the English conjunction, with an interrogative "pedigree" and with a broader set of subsidiary syntactic functions. The verb 'to know' is the immediate governor of clause (2). The clause could have the pronoun

yah ‘it’, ‘this’ in the position before the infinitive ‘know’ serving as a cataphoric prop for the subordinate clause (2).

The verb ‘to know’ governing clause (2) is, in its turn, the object complement of the finite predicate ‘want’. This verb is sub-classified as a verb of mental attitude, a semantic sub-category with a range of idiosyncratic features, the capacity to take a clausal complement along with one or two nominal arguments (Pearson 2021). The nominal arguments may be of predicative nature: infinite verbal forms extending the combining capacity of the clause. The bi-verbal character of the finite VP in the main clause determines its double syntactic valency, that is its capacity to subordinate two clauses — (2) and (3). The former has one infinite verb form as its governor, whereas the syntactic governor of the latter is the whole bi-verbal VP. The semantic scope of the clause (3) dependency is still broader—it is the whole complex sentence. The connecting device *kyōki* ‘because’ (*kyō-ki* ‘why-[subordinator] that’) is a fully-fledged conjunction, as is to be expected in adverbial clauses (cf. Lehmann 1988). The dependency distance of the adverbial clause is longer than that of the object clause. This also points to its rather loose formal connection within the string, which is obviously compensated for semantically.

A more interesting, although predictable, aspect of the syntactic asymmetry is the inverse proportionality between the strength of the syntactic connection and the semantic dependency scope: the object complement clause, which is firmly embedded in the main clause, is semantically connected to its predicate, whereas the juxtaposed adverbial clause is semantically linked to the whole preceding complex sentence.

In the following excerpt the structures incorporating clauses (4) to (7) can be ignored in the discussion (hence they are marked with double slashes on both sides). They help us to understand the context of the sequence (8) to (11) which builds to a longer string—string (B).

//*Yah sunkar atithi muskurāe aur bole₄ – hamāre āpke bīc Bādrāyaṅ sambamdh hai₅.*
 this having.heard guest smiled and said₄ – me you between Bādrāyaṅ’s⁵ relation is₅.
Jaise āpke darvāze par ber (badrī phal) kā peṛ hai₆, vaise hī mere darvāze par bhī ber
 Like your door at jujube (badrī fruit) of tree is₆, so exactly my door at too jujube
kā peṛ hai₇.
 GEN tree is₇.//

(B) *cūki rāt ho rahī thī₈ aur pūrā gāv mere lie aparicit thā₉, islie*
 as night falling was₈ and whole village me for unknown was₉, therefore
maīne yah sambamdh nikālā₁₀ ki kuch to samāntā hai mere āp=mē₁₁.
 I this relation thought.up₁₀ that some at.least similarity is me you=in₁₁.

5 The idiomatic expression ‘Bādarāyaṅ’s relation’ means ‘a far-fetched, just a nominal relation’.

(<CxS {cause-eff} {cml} > [(<CxS {cause-eff} > [< {cũki} SC₈ > < {aur} < {cũki} > SC₉ > < {islie} MC_{10i} :val1 >] [(<CxS {cml} > [< MC_{10i} > :val2 > < {ki} SC₁₁ >]]]]))

//‘Having heard this the guest smiled and said₄, “We are distantly related₅. Just as there is a jujube tree in front of your house₆, so there is also a jujube tree in front of my house₇.”’//

(B) ‘As night was falling₈ and I didn’t know anybody in the village₉, I therefore thought up this relationship₁₀, so that at least there is something in common between us₁₁.’

String (B) is a complex sentence. It consists of four units with subordinate and coordinate interclausal relations. Although it includes both relations types it is not considered a combined sentence because the coordinate link is located not on the highest level of the string. Subordinate relations prevail in (B), whereas the only coordinate bond expressed by the conjunction *aur* ‘and’ connects two collateral subordinate clauses (8) and (9). They build a sequence and share the causal conjunction *cũki* ‘as’, which connects them to the nucleus⁶ of the string—the main clause (10) introduced by the adverbial connector *islie* ‘therefore’.

All clauses are introduced by connectors: three of them by conjunctions—*cũki* ‘as’ (8), *cũki* ‘as’... *aur* ‘and’ (9), *ki* ‘that’ (11)—and one by the adverbial connector *islie* ‘therefore’ (10). The conjunctions *cũki* and *ki* are borrowings from Persian. Both are functional derivations of interrogative-relative pronouns. The former is a combination of *cũ* ‘where’, ‘why’, ‘how’ and *ki* (< ‘who’, ‘what’), which can introduce almost any subordinate clause and which appears as a separate conjunction in the final clause of the sentence. The coordinating conjunction *aur* (< Skt. *apara-* ‘other’, see Turner 1966: 20) conjoins two causal subordinate clauses. The connector *islie* gives the clause the meaning of effect and marks it as the main clause. It consists of the oblique form of the deictic pronoun *is* ‘it’ and the deverbal marker *lie* (< *lenā* ‘to take’), the meaning on the whole being approximately ‘this taken’. The pronominal anaphor refers to the situation rendered by the preceding proposition of the subordinate constituent. Here we have a case where each part of the cause-effect relation is marked with its own device making explicit the meaning of the interclausal link. However, this tandem is not absolutely necessary and some language purists even consider the double marking stylistic negligence. Each of the markers alone serves the same semantic effect, but the syntactic connection is then realised differently in each case. If the causal

6 The term ‘nucleus’ is understood here as the sentence core, not in the sense in which it is used by Longacre (1970; 1985) and in subsequent studies adopting the same or a similar approach.

marker *cūki* has zero correspondence in the effect clause, which is the main clause, the clauses build a sentence. If there is explicit marking only in the effect clause, the syntactic integrity is weakened and both clauses may be considered separate sentences. According to my preliminary observations, the latter way of mapping the causal relation is the most frequent one among several marked types of Hindi multiclausal causal constructions.

The second valency of the main (effect) clause is induced by the object of the finite verb: *sambamdh* ‘relation’. It is determined by the descriptive relative clause with the conjunction *ki* ‘that’ specifying the noun and correlated with the preposed deictic pronominal attribute *yah* ‘this’.

- (C) *yah sunkar sabne zor kā thahākā lagāyā aur atithi ko*
 This having.heard all strength of laughter laid.out and guest OBJ
sādar vidā kiyā, yah kahkar₁₂ ki₁₃ jab sambamdh sthāpit
 respectfully see.off did, this having.said₁₂ that_{13#} when relation established
ho hī gayā hai₁₄, to_{15#} jab kabhī idhar se guzrē₁₆,
 be really gone is₁₄, then_{15#} when sometime here through would.pass.by₁₆,
yahī rātri-vīsrām karē_{15;13}.
 here.only night-rest would.do_{15;13}.

(<CxS {cmpl} {cause-eff} {tmp}>[<MC₁₂-L4<<{ki}SC₁₃-L3>](<CxS {cause-eff}>
 [{jab/cause}SC₁₄-L2<<{to}MC₁₅-L3:val1>](CxS {tmp}>[<{jab}SC₁₆-L1>
 <<MC₁₅-L3:val2 >>]))))

‘Having heard this [they] all guffawed and bade the guest a respectful farewell saying₁₂ that_{13#}, “As the relationship has been established₁₄, so_{15#} whenever you pass by here₁₆, stay only here [in this house] for the night₁₅”₁₃.’

String C is a four-level complex sentence with stepwise subordination. The fourth level main clause (12) joins the postposed subordinate clause (13–16) due to the valency of a verb of saying (used in the converb form *kahkar* ‘having said’). With regard to its syntactic structure the subordinate clause is a tripartite complex sentence conjoined to the fourth-level main clause by the complementiser *ki*. Its main clause (15), located on the third level, has two valences. One of them is filled by a preposed clause (14) and the other one by an embedded clause (16). Both clauses are introduced by the conjunction *jab* ‘when’ but they differ in relation to the main clause (15) with regard to both semantics and syntax. In the subordinate clause (14) *jab* is used as a cause marker and correlated with the conjunction *to* introducing the main clause. (The conjunction is marked as a distant clause part.) In Hindi, markers of main clauses are generally more significant in establishing interclausal relations than markers of subordinate

clauses; as a rule, these can easily be omitted. The embedded subordinate clause (16) uses the time conjunction in its basic meaning of time. This marker has no correlative in the main clause of the complex sentence, as an embedded clause does not need any further tie to the matrix clause. The connection is clear from the clause location and the subordinating conjunction *jab*. Thus, two subordinate clauses occupy different levels: (14) is on the second level, whereas (16) demonstrating the strongest bond with the main clause (15) is on the first level. The positions of the subordinate clauses (14) and (16) in the syntactic hierarchy of the string may be correlated with their semantics. It has been suggested that causal clauses' connection to the main clause is the loosest among all other semantic types of adverbial clauses (Diessel & Hetterle 2011). It is conceivable that adverbial clauses with the basic adverbial meanings of time and space, which are expressed in clauses by lexical adverbial modifiers, enjoy a closer relation to the main clause.

- (D) *saṃyog se, us din khetō mē kām zyādā thā₁₇, aur koī sarkārī*
 Chance by that day fields in work much was₁₇ and any government
naukrī tō thī nahī₁₈ ki kām pūrā ho na ho₁₉, gharī
 job indeed was not₁₈ that work finished be.SBJV not be.SBJV₁₉ clock
dēkhkar log ghar bhāgē₂₀. so, unke āte-āte kāfī der
 look.CVB people home run₂₀. So their coming-coming enough tardiness
ho gāī₂₁.
 be went₂₁.

String{cause-eff}(<CdS>([<CpS{aur}>[<C₁₇-L3>(<{aur}>CxS{attr}>[<MC₁₈-L3>
 <{ki}>SC₂₀-L2>])<CxS{cnecs}>(<SC₁₉-L1><MC₂₀-L2>]<{so/eff}>MC₂₁-L4>))))

‘By chance that day there was much work in the fields₁₇, and it wasn’t a government job₁₈, in which whether the work is finished or not₁₉, the people look at the clock and run home₂₀. So when they [the menfolk of the family] came it was [already] late₂₁.’

Unlike the clause strings (A)–(C), string (D) does not fit into any definition of a sentence. In establishing the boundaries of clause strings which do not fit into any definition of a sentence this study draws on the semantic interrelations characterising a multiple event sequence as the ground for delimitation of strings.

(D) is a five-part string which includes two syntactic segments graphically framed as sentences. The concluding simple sentence is separated from the preceding combined sentence of four clauses by the Devanagari full stop sign. Nevertheless, it belongs semantically to the same string, presenting the final event of the whole situation and building a clear logical transition to the following complex

situation. The fifth constituent is introduced by the pronominal conjunction *so* ‘so’, which may introduce a new graphical sentence, as in this case. In similar contexts it may be separated from the previous part of a sentence by a comma or not separated at all (the same as in English).⁷ This variation in punctuation shows the possible transitions between syntactic independence and a tighter formal bond to the semantic correlate within a clausal string. This kind of alternating syntactic framing seems to be typical of consecutive clauses.

The string has a four-level hierarchical structure. At its highest point (fourth level) is the final clause (21), which is related to the whole preceding clause sequence as effect to cause. The third level is formed by a simple (17) and a complex (18–20) sentence connected by the coordinating conjunction *aur*. The complex sentence exhibits stepwise subordination forming the second level: the rightward valency of the main clause (18) is filled by a continuative relative clause (19–20), which depends on the main clause subject and expands its content. Finally, the unmarked conditional concessive relation between the subordinate (19) and the main (20) clause is located on the lowest (first) level.

6 In place of conclusion. Future directions of the data analysis and formalisation: a view

The intended study needs a variety of numerically reliable information in the form of a database. Such a database built on written and oral sources will be useful for relational research in Hindi linguistics and in typology. The generally estimated workable database volume is set at 5 million words with a desirable (but in our case unrealistic) expansion up to a sample of 20 million words (Matthiessen 2002: 252). Currently, the primal data universe comprises somewhat more than 100 data units.

The perspective on the general issues needs to be broadened to cover data elicited not only from written texts but also oral discourse. Inclusion in large quantities of discourse material in data will shed a new light on Hindi complex communication structures and build a solid basis for exploring the cognitive characteristics of its syntactic complexity.

Complex syntax is tightly related to variation. The more complex the situation, the more propositions participate in its portrayal. Accordingly, the number of clauses framing them is higher which allows a higher variation in their linkage. This in turn leads to an increase in variation in perspectives on the situation. Variance in

7 Compare: *gāv̄ ke sāre log aśikṣit haī, so mazdūrī par zimdā rahte haī, [...]* ‘All villagers are uneducated, so they live from their manual labour, [...]’ (Simh 2018) and *ārthik sthiti thīk nahī so mazdūrī kī majbūrī hai*. ‘[They are] not well financially so manual work is a necessity.’ (“Dūr nahī...”) / *Dainik Bhāskar*.

the interplay between meanings and formal tools brings up the issue of “choice” as propounded by Halliday (Halliday & Matthiessen 2014), in the sense of a speaker’s primary choice of meaning and its correlation with multiple forms.

Processing original Hindi texts aims, among other things, at compiling count data on the total number of clauses, in particular, of clauses with adverbial meanings, as well as on variations in them. These have to do with the semantics of the interclausal bonds. Further variations have to do with the information structure and can impact the placing of conjunctions. The feature [+/- focused] can also be responsible for the clausal order. Such phenomena as foregroundedness and backgroundedness also come into play.

A productive and, in terms of text generation, promising analysis would involve contrasting bipartite complex components. Two types of opposition between ways of framing interclausal adverbial relations should be considered: 1. various adverbial relations within the same clause string and 2. same adverbial relations in different clause strings. The major analytic perspective would start from the semantic vantage point and integrate contextually conditioned choices of structural-syntactic and auxiliary lexico-grammatical means of expression.

In terms of information structure, the relation between parts of a minimal—bipartite—hypotactic string is interpreted in the following way: a governing/main clause profiles a process that is foregrounded, while the sub-clause profiles a backgrounded process, be it causal, conditional or circumstantial. Various types of conceptual and functional subordination underlie the subordinating structures, which are shaped with the help of grammatical and lexico-grammatical means.

The exploration is underpinned by the general observation that some types of intersentential relations may be explicated formally or else be expressed without using any special formal means. The latter type of connection results from the order of sentences and the semantics of their key terms; the former type uses semantically specific markers.

Complex sentences in Hindi belong to the marked type. The markers in a number of cases may be dependent on the meaning of the interclausal relation modified by the content of the protasis and apodosis. A large formalised database should provide a reliable foundation for establishing the marking rate, positional and scope variations of the markers as well as semantic and logical grounds of their overlapping. Further tasks deal with information structure and communicative functions of the formal varieties (structuring types of composite sentences) and amorphous types of clausal linkage.

The part of the database planned to formalise oral data offers much more significant challenges on each step of the process—from collecting data to parsing the speech stream and stratifying the relations between the syntactic segments. It is axiomatic that dia- and polylogues present the most challenging problem. This

lies in the unpredictable turns a syntactic trajectory may take at any moment in the time resulting in clausal structures within the speaker's turn domain or across the speakers' turns (Auer 2005).

Higher syntax in Hindi still awaits a thorough exploration of its written and especially oral discourse incarnations. The formalised corpus is expected to eventually provide a substantial basis for an insightful conceptualisation of clause ordering, text structure and the varied relationships between semantics of inter-clausal connections and the language means marking them. Enhanced data and new foci, especially a cognitive one, will bring a new dimension also to typological research involving Hindi. The results of the future work can be used to enhance modelling of probabilistic sequences of clauses and sentences, also in descriptive, analytic and computational linguistics.

ORCID®

Tatiana Oranskaia  <https://orcid.org/0000-0001-7207-9190>

Bibliography

- Aguiar, Joana & Pilar Barbosa 2016. "Establishing Causal Relations in European Portuguese: Coordination and Subordination in a Stratified Corpus", in: Pratas, Fernanda, Sandra Pereira and Clara Pinto (eds): *Coordination and Subordination: Form and Meaning—Selected Papers from CSI Lisbon 2014*. Newcastle upon Tyne: Cambridge Scholars Publishing, 1–22.
- Ananthanarayana, H. S. 1996. "Relative Clause in Indo-Aryan", in: Swarajya, Lakshmi & Aditi Mukherjee (eds): *Word Order in Indian Languages*. Hyderabad: Centre of Advanced Study in Linguistics, Osmania University, Hyderabad; Booklinks Corporation, 31–46.
- Auer, Peter 1992. "The Neverending Sentence: Rightward Expansion in Spoken Language", in: Kontra, Miklós (ed.): *Studies in Spoken Languages: English, German, Finno-Ugric*. (Sonderdrucke aus der Albert-Ludwigs-Universität Freiburg. Budapest: Hungarian Academy of Sciences, Linguistics Inst.), 41–59.
- Auer, Peter 2005. "Projection in Interaction and Projection in Grammar", *Text & Talk. An Interdisciplinary Journal of Language, Discourse & Communication Studies* 25,1: 7–36. (<<https://doi.org/10.1515/text.2005.25.1.7>>, accessed: September 6, 2020).
- Auer, Peter et al. 2009. "Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)", *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 10: 353–402 (<<http://www.gespraechsforschung-ozs.de>>, accessed: February 1, 2021).

- Bhatt, Rajesh 2003. “Locality in Correlatives”, *Natural Language & Linguistic Theory* 21,3: 485–541. (<<http://www.jstor.org/stable/4048041>>, accessed: March 5, 2021).
- Bhat Riyaz Ahmad et al. 2017. “The Hindi/Urdu Treebank Project”, in: Ide, Nancy & James Pustejovsky (eds): *Handbook of Linguistic Annotation*. Dordrecht: Springer (<https://doi.org/10.1007/978-94-024-0881-2_24>, accessed: September 12, 2020)
- Bonial, Claire et al. 2015. *English PropBank Annotation Guidelines*. Center for Computational Language and Education, Research Institute of Cognitive Science, University of Colorado at Boulder. (<<http://verbs.colorado.edu/propbank/EPB-Annotation-Guidelines.pdf>>, accessed: May 5, 2019).
- Chafe, Wallace 1984. “How People Use Adverbial Clauses”, in: *Proceedings of the Tenth Annual Meeting of the Berkeley Linguistics Society*, 437–449. (<<http://journals.linguisticsociety.org/proceedings/index.php/BLS/article/viewFile/1936/1708>>, accessed: November 13, 2019).
- Christofaro, Sonia 2014. “Is There Really a Syntactic Category of Subordination?”, in: Visapää, Laura, Jyrki Kalliokoski & Helena Sorva (eds): *Contexts of Subordination: Cognitive, Typological and Discourse Perspectives*. (Pragmatics & Beyond New Series, vol. 249.) Amsterdam; Philadelphia: John Benjamins Publishing Company, 73–91.
- Dayal, Veneeta 1996. *Locality in WH Quantification: Questions and Relative Clauses in Hindi*. (Studies in Linguistics and Philosophy 62.) Dordrecht; London: Kluwer.
- Diessel, Holger 2005. “Competing Motivations for the Ordering of Main and Adverbial Clauses”, *Linguistics* 43,3: 449–470. (<<https://doi.org/10.1515/ling.2005.43.3.449>>, accessed: February 5, 2018).
- Diessel, Holger & Katja Hetterle 2011. “Causal Clauses: A Crosslinguistic Investigation of Their Structure, Meaning, and Use”, in: Siemund, Peter (ed.): *Linguistic Universals and Language Variation*. (Trends in Linguistics. Studies and Monographs 231.), 23–54.
- “Dūr nahī – rājdhānī Patnā se Sipārā pul ke nīce roz”, *Dainik Bhāskar* (date cannot be retrieved) (<<https://www.bhaskar.com/latest-patna-news-043521-2483803.html>>, accessed: March 29, 2021).
- Eifring, Halvor 1995 [1993]. *Clause Combination in Chinese*. (Sinica Leidensia 32.) Leiden; New Yourk; Köln: E. J. Brill.
- Ernst, Thomas 2020. “The Syntax of Adverbials”, *Annual Review of Linguistics* 6,1: 89–109 (<<https://www.annualreviews.org/doi/10.1146/annurev-linguistics-011619-030334>> accessed: March 15, 2021).

- Fanselow, Gisbert, Caroline Féry, Matthias Schlesewsky & Ralf Vogel (eds) 2006. *Gradience in Grammar: Generative Perspectives*. Oxford; New York: Oxford University Press.
- Foley, William A. & Van Valin, Robert D., Jr. 1984. *Functional Syntax and Universal Grammar*. Cambridge: Cambridge University Press.
- Givón, Talmy 1990. *Syntax. A Functional-Typological Introduction*, vol. II. Amsterdam: John Benjamins.
- Haiman, John & Sandra A. Thompson (eds) 1988. *Clause Combining in Grammar and Discourse*. (Typological Studies in Language 18.) Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Halliday, Michael A. K. 1985. *An Introduction to Functional Grammar*. London: Arnold.
- Halliday, Michael A. K. & Christian Matthiessen 2014 [1985; 1994; 2004]. *Halliday's Introduction to Functional Grammar*. New York: Routledge.
- Haspelmath, Martin 2004. "Coordinating Constructions: An Overview", in: Haspelmath, Martin (ed.): *Coordinating Constructions*. Amsterdam; Philadelphia: John Benjamins Publishing Company, 3–39.
- Heine, Bernd, Gunther Kaltenböck & Tania Kuteva 2016. "On Insubordination and Cooptation", in: Watanabe, Honoré & Nicholas Evans (eds): *Insubordination*. Amsterdam; Philadelphia: John Benjamins, 39–64.
- Herlin, Ilona, Jyrki Kalliokoski, and Lana Visapää 2014. "Introduction", in: Visapää, Laura, Jyrki Kalliokoski, and Helena Sorva (eds): *Contexts of Subordination: Cognitive, Typological and Discourse Perspectives* (Pragmatics & Beyond New Series, vol. 249.) Amsterdam; Philadelphia: John Benjamins Publishing Company, 1–16.
- Ide, Nancy M. & Laurent Romary 2004. "International Standard for a Linguistic Annotation Framework", *Natural Language Engineering* 10,3–4: 211–225. (<<https://doi.org/10.1017/S135132490400350X>>, accessed: January 9, 2017).
- Jucker, Andreas H. 1991. "Between Hypotaxis and Parataxis. Clauses of Reason in Ancrène Wisse", in: Kastovsky, Dieter (ed.): *Historical English Syntax*. (Topics in English Linguistics.) Berlin; New York: Mouton de Gruyter, 203–220.
- Kachru, Yamuna 1978. "On Relative Clause Formation in Hindi-Urdu", *Linguistics* 207: 5–26.
- Kortmann, Bernd 1997. *Adverbial Subordination: A Typology and History of Adverbial Subordinators Based on European Languages*. (Empirical Approaches to Language Typology 18.) Berlin; New York: Mouton de Gruyter.
- Kothari, Anubha 2010. *Processing Constraints and Word Order Variation in Hindi Relative Clauses*. Stanford University, ProQuest Dissertations Publishing, 28168220. (<<http://purl.stanford.edu/vh402hk4761>>, accessed: October 17, 2020).

- Langacker, Robert 2010 [1998]. “Conceptualization, Symbolization and Grammar”, *International Journal of Cognitive Linguistics* 1: 31–63.
- Langacker, Ronald W. 2014. “Subordination in a Dynamic Account of Grammar”, in: Visapää, Laura, Jyrki Kalliokoski & Helena Sorva (eds). *Contexts of Subordination: Cognitive, Typological and Discourse Perspectives* (Pragmatics & Beyond New Series, vol. 249). Amsterdam; Philadelphia: John Benjamins Publishing Company, 17–72.
- Ledgeway, Adam 2016. “From Coordination to Subordination: the Grammaticalisation of Progressive and Andative Aspect in the Dialects of Salento”, in: Pratas, Fernanda, Sandra Pereira & Clara Pinto (eds): *Coordination and Subordination. Form and Meaning—Selected Papers from CSI Lisbon 2014*. Newcastle upon Tyne: Cambridge Scholars Publishing, 157–184.
- Lehmann, Christian 1985. “On Grammatical Relationality”, *Folia Linguistica* 19: 67–109. (<<https://www.degruyter.com/document/doi/10.1515/flin.1985.19.1-2.67/html>>, accessed: October 17, 2020).
- Lehmann, Christian 1988. “Towards a Typology of Clause-Linkage”, in: Haimann, John & Sandra Thompson (eds): *Clause Combining in Grammar and Discourse*. Amsterdam; Philadelphia: John Benjamins Publishing Company, 181–225. (<<https://www.christianlehmann.eu/publ/linkage.pdf>>, accessed: July 15, 2020)
- Longacre, Robert E. 1960. “String Constituent Analysis”, *Language* 36,1(1): 63–88.
- Longacre, Robert E. 1970. “Sentence Structure as a Statement Calculus”, *Language* 46,4: 783–815.
- Longacre, Robert E. 2007 [1985]. “Sentences as Combinations of Clauses”, in: Shopen, Timothy (ed.): *Language Typology and Syntactic Description*. Vol. II: *Complex Constructions*. Cambridge: Cambridge University Press.
- Matthiessen, Christian M. I.M. 2002. “Combining Clauses into Clause Complexes: A Multi-faceted View”, in: Bybee, Joan L. & Michael Noonan (eds): *Complex Sentences in Grammar and Discourse: Essays in Honor of Sandra A. Thompson*. Amsterdam; Philadelphia: John Benjamins Publishing Company, 235–320.
- Miśra, Buddhināth s. a. *Phūl āe haī kanerō mē* (Oleanders are Blossoming), (<<http://www.hindisamay.com>>, accessed: January 20, 2018).
- Oranskaya, Tatyana I. 2005. “Conditional Constructions in Hindi”, in: Xrakovskij, Victor S. (ed.): *Typology of Conditional Constructions*. (LINCOM Studies in Theoretical Linguistics 25.) Munich: Lincom Europa, 218–245.
- Pearson, Hazel 2021. “Attitude Verbs”, in: Gutzmann, Daniel et al. (eds): *Companion to Semantics*. Hoboken: Wiley Blackwell, 2021. (<<https://semanticsarchive.net/Archive/mIwMTc1Z/Attitude%20verbs%20Pearson%20to%20appear.pdf>>, accessed: January 30, 2021).

- Ricca, Davide 2010. “Adverbs”, in: Baldi, Philip & Pierluigi Cuzzolin (eds): *New Perspectives in Historical Latin Syntax*. Vol. 2. *Constituent Syntax: Adverbial Phrases, Adverbs, Mood, Tense*. (Trends in Linguistics. Studies and Monographs 180,2.) Berlin; New York: De Gruyter Mouton, 109–191.
- Ross, Daniel 2016. “Between Coordination and Subordination: Typological, Structural and Diachronic Perspectives on Pseudocoordination”, in: Pratas, Fernanda, Sandra Pereira & Clara Pinto (eds): *Coordination and Subordination. Form and Meaning—Selected Papers from CSI Lisbon 2014*. Newcastle upon Tyne: Cambridge Scholars Publishing, 209–244.
- Sharma, Ghanshyam 2010. “On Hindi Conditionals”, in: *Annual Review of South Asian Languages and Linguistics*, 107–134.
- Sharma, Ghanshyam 2012. “On the Role of Protases in Conditional Statements: Some Evidence from Hindi”, in: *Annual Review of South Asian Languages and Linguistics 2011*, 49–78.
- Shopen, Timothy (ed.) 2007 [1985]. *Language Typology and Syntactic Description*. Vol. 2: *Complex Constructions*. Cambridge: Cambridge University Press.
- Simh, Tavlin 23 Jan 2018. Dalitō kī āvāz bhī sunnī paregī. *Amar Ujālā* (<<https://www.amarujala.com/columns/opinion/the-voice-of-the-dalits-will-also-be-heard>>, accessed: March 29, 2021).
- Thompson, Sandra A., Robert E. Longacre & Shin Ja J. Hwang 2007 [1985]. “Adverbial Clauses”, in: Shopen, Timothy (ed.): *Language Typology and Syntactic Description*. Vol. 2. *Complex Constructions*. Cambridge: Cambridge University Press, 237–300.
- Traugott, Elizabeth C. & Graeme Trousdale (eds) 2010. *Gradience, Gradualness & Grammaticalization*. Amsterdam; Philadelphia: John Benjamins Publishing Company.
- Umangi, Oza et al. 2009. “Experiments with Annotating Discourse Relations in the Hindi Discourse Relation Bank”. (<<https://www.researchgate.net/publication/228574312>> accessed: April 11, 2020).
- Wang, Yaqin & Jianwei Jan 2018. “A Quantitative Analysis on a Literary Genre Essay’s Syntactic Features”, in: Jiang, Jingyang & Haitao Liu (eds): *Quantitative Analysis of Dependency Structures*. (Quantitative Linguistics 72.) Berlin; Boston: De Gruyter Mouton, 295–314. (<https://www.researchgate.net/profile/Yaqin-Wang-3/publication/328147823_A_Quantitative_Analysis_on_a_Literary_Genre_Essay%27s_Syntactic_Features/links/5d6a0f20a6fdcc547d6d195e/A-Quantitative-Analysis-on-a-Literary-Genre-Essays-Syntactic-Features.pdf>, accessed: October 8, 2018)