Journal of Language Modelling

VOLUME 9 ISSUE 2
DECEMBER 2021

Institute of Computer Science
Polish Academy of Sciences
Warsaw

Journal of Language Modelling

VOLUME 9 ISSUE 2
DECEMBER 2021

Articles

JOURNAL OF
LANGUAGE MODELLING

ISSN 2299-8470 (electronic version)
ISSN 2299-856X (printed version)
http://jlm.ipipan.waw.pl/

MANAGING EDITOR
Adam Przepiórkowski IPI PAN

SECTION EDITORS
Elżbieta Hajnicz IPI PAN
Agnieszka Mykowiecka IPI PAN
Marcin Woliński IPI PAN

STATISTICS EDITOR
Łukasz Dębowski IPI PAN

Published by IPI PAN
Institute of Computer Science, Polish Academy of Sciences
ul. Jana Kazimierza 5, 01-248 Warszawa, Poland

Circulation: 50 + print on demand

Layout designed by Adam Twardoch.
Typeset in XƎLaTeX using the typefaces: *Playfair*
by Claus Eggers Sørensen, *Charis SIL* by SIL International,
JLM monogram by Łukasz Dziedzic.

All content is licensed under
the Creative Commons Attribution 4.0 International License.
http://creativecommons.org/licenses/by/4.0/

EDITORIAL BOARD

Steven Abney University of Michigan, USA

Ash Asudeh Carleton University, CANADA;
University of Oxford, UNITED KINGDOM

Chris Biemann Technische Universität Darmstadt, GERMANY

Igor Boguslavsky Technical University of Madrid, SPAIN;
Institute for Information Transmission Problems,
Russian Academy of Sciences, Moscow, RUSSIA

António Branco University of Lisbon, PORTUGAL

David Chiang University of Southern California, Los Angeles, USA

Greville Corbett University of Surrey, UNITED KINGDOM

Dan Cristea University of Iași, ROMANIA

Jan Daciuk Gdańsk University of Technology, POLAND

Mary Dalrymple University of Oxford, UNITED KINGDOM

Darja Fišer University of Ljubljana, SLOVENIA

Anette Frank Universität Heidelberg, GERMANY

Claire Gardent CNRS/LORIA, Nancy, FRANCE

Jonathan Ginzburg Université Paris-Diderot, FRANCE

Stefan Th. Gries University of California, Santa Barbara, USA

Heiki-Jaan Kaalep University of Tartu, ESTONIA

Laura Kallmeyer Heinrich-Heine-Universität Düsseldorf, GERMANY

Jong-Bok Kim Kyung Hee University, Seoul, KOREA

Kimmo Koskenniemi University of Helsinki, FINLAND

Jonas Kuhn Universität Stuttgart, GERMANY

Alessandro Lenci University of Pisa, ITALY

Ján Mačutek Comenius University in Bratislava, SLOVAKIA

Igor Mel'čuk University of Montreal, CANADA

Glyn Morrill Technical University of Catalonia, Barcelona, SPAIN

Stefan Müller Freie Universität Berlin, GERMANY

Mark-Jan Nederhof University of St Andrews, UNITED KINGDOM

Petya Osenova Sofia University, BULGARIA

David Pesetsky Massachusetts Institute of Technology, USA

Maciej Piasecki Wrocław University of Technology, POLAND

Christopher Potts Stanford University, USA

Louisa Sadler University of Essex, UNITED KINGDOM

Agata Savary Université François Rabelais Tours, FRANCE

Sabine Schulte im Walde Universität Stuttgart, GERMANY

Stuart M. Shieber Harvard University, USA

Mark Steedman University of Edinburgh, UNITED KINGDOM

Stan Szpakowicz School of Electrical Engineering and Computer Science, University of Ottawa, CANADA

Shravan Vasishth Universität Potsdam, GERMANY

Zygmunt Vetulani Adam Mickiewicz University, Poznań, POLAND

Aline Villavicencio Federal University of Rio Grande do Sul, Porto Alegre, BRAZIL

Veronika Vincze University of Szeged, HUNGARY

Yorick Wilks Florida Institute of Human and Machine Cognition, USA

Shuly Wintner University of Haifa, ISRAEL

Zdeněk Žabokrtský Charles University in Prague, CZECH REPUBLIC

Aspects of linguistic ageing in literary authors across time

Carmen Klaussner, Carl Vogel, and Arnab Bhattacharya
Trinity College Dublin

ABSTRACT

This work offers an investigation into linguistic changes in a corpus of literary authors hypothesized to be attributable to the effects of ageing. In part, the analysis replicates an earlier study into these effects, but adds to it by explicitly analyzing and modelling competing factors, specifically the influence of background language change. Our results suggest that it is likely that this underlying change in language usage is the primary force for the change observed in the linguistic variables that was previously attributed to linguistic ageing. However, our results are tentative insofar as we do not examine non-linear models in general, or other variables influenced by ageing, or non-professional writers who may be more susceptible to these observed shifts in general language than was observable for the literary authors.

Keywords:
linguistic ageing,
diachronic literary analysis,
language change

INTRODUCTION 1

Language is subject to constant change, both with respect to a particular linguistic variety that affects all its speakers as well as on an individual level for each speaker separately during their lifetime. "Stylometry" is the study of a writer's stylistic fingerprint based on collected writings over his or her lifetime. Due to the sequential and long term nature of publishing, stylometric studies may be influenced by temporal language development and its effects might be misconstrued and misinterpreted as a result. More recently this issue has given rise to a temporal variant of stylometric analysis, i.e. "stylochronometry", that studies changes in style over time, as exemplified, for instance, by the work of Forsyth (1999), Stamou (2007) and Klaussner and Vogel (2018b). However, even though stylochronometric studies consider the temporal dimension, these analyses still conflate individual stylistic changes with those induced by ageing, such as changes in authors' vocabulary size over time and, most importantly, influences that affect speakers of the same language variety equally, such as general underlying language shifts. Although previous studies have examined sets of linguistic variables with respect to both healthy and pathological ageing (Pennebaker and Stone 2003; Le *et al.* 2011; Kemper *et al.* 2001), to the best of our knowledge there do not yet exist composite studies considering all three aforementioned factors.

The current work extends a research paradigm created by Pennebaker and Stone (2003) who analyzed linguistic ageing both in emotional disclosure studies and in a corpus of literary authors. In this work, we build on previous results by examining a larger literary corpus as well as controlling for background language change. Our objective is to replicate the earlier study on a different literary corpus that is temporally-aligned with a reference corpus for that same time period, thus allowing us to investigate possible influences of general language shifts. We also propose some methods that can be used to attempt to disentangle general effects from those that are individual.

Within this paper, Section 2 discusses previous work in the area, Section 3 presents the literary authors data set, Section 4 and Section 5 discuss methods and experiments respectively and Section 6 and Section 7 analyze and summarize the results.

2 RELATED WORK

Patterns of general, underlying language change have been studied by, for instance Lieberman *et al.* (2007), finding that the question of whether an irregular verb in English will acquire the "-ed" regularization largely depends on its token frequency. Highly entrenched, irregular verbs such as *have* or *be* are less likely to be regularized. Štajner and Mitkov (2011) investigated diachronic changes in American (AE) and British English (BE) with respect to four different variables: Average Sentence Length (ASL), Automated Readability Index (ARI), Lexical Density (LD) and Lexical Richness (LR) across four different text categories (press, general prose, learned and fiction)[1,2] Based on two-tailed t-tests, they report a statistically different increase for ARI in BE press/prose (interpreted by the authors as a tendency to render texts more difficult to read in these categories), while the ASL for BE did not change significantly in the period of 1961–1991. Both LR and LD increased across each of the press, prose and fiction categories. In comparison, while AE does not exhibit a significant change in ARI, ASL decreased significantly for the press and learned text categories, which is interpreted as an example of colloquialization. LR and LD in AE only increased in the prose text category. Statistically significant differences between 1961 AE press and 1961 BE press for ARI/LD/LR disappeared by 1991/1992 which is attributed to the growing Americanization that would be particularly tangible in this category.

One of the first statistically-oriented studies into changes in an author's writing style was Forsyth's (1999) study of the poet W. B. Yeats. The study's objectives were to develop stable methods for chronological prediction as well as to examine possible changes in Yeats' style, the exact manifestation of which is disputed among literary scholars. The analysis considered distinctive marker substrings extracted from 142 poems using a modified version of "Monte-Carlo Feature Finding" (a quasi-random search algorithm). Features were then ranked according to distinctiveness measured by χ^2 in separating the categories "Young Yeats" (before 1915) and "Old Yeats" (after 1915). Forsyth (1999) reported identifying clear markers of young and old Yeats based on 20 substring markers: for nine out of ten test poems their count is higher in the appropriate age category.

Another literarily-motivated analysis (Hoover 2007) considered the late 19th century American author Henry James, who supposedly

[1] Published: 1961 (BE + AE) and 1991 (BE)/1992 (AE).

[2] Lexical density is defined as 'number of unique tokens/total number of tokens', whereas lexical richness is defined as 'number of unique lemmas/total number of tokens'.

changed his style over his creative lifespan. Based on literary scholars' findings (e.g. Beach 1918), Hoover investigates natural partitions of James' style into three different temporal divisions of early (1877–1881), intermediate (1886–1890) and late style (1897–1917) using the most frequent word unigrams and a variety of different methods, such as Cluster Analysis, Burrows' Delta, Principal Component Analysis and Distinctiveness Ratio.[3] Apart from these divisions, Hoover also notes the existence of gradual transitions in between, with for instance the first novels of the late period being somewhat different from the rest of them.

Le *et al.* (2011) contrasted the writings of three female British novelists for detecting markers of dementia, specifically Iris Murdoch, who died with Alzheimer's disease, Agatha Christie, who was suspected of having it, and P. D. James, who aged healthily. Previous research (Kemper *et al.* 2001; Bird *et al.* 2000; Burke and Shafto 2008 as cited by Le *et al.* 2011) indicated that for instance vocabulary and syntactic complexity declined more rapidly in the presence of dementia, particularly with respect to words of lower frequency and higher specificity as well as passive constructions. Simultaneously, occurrence of lexical repetitions and disfluencies would increase. Analyzing a variety of lexical and syntactic measures, Le *et al.* (2011) could largely confirm their hypotheses with regard to more rapid lexical decline in Murdoch. More than 20 years before any Alzheimer's symptoms became apparent, her vocabulary started to decline, resulting in a significant increase in lexical repetitions of content words. However, her lexical specificity, measured through the proportion of specific indefinite nouns and verbs, remained intact throughout. All but two of Christie's lexical types showed an overall decline. In contrast, the vocabulary, repetition and specificity scores vary only slightly across James' novels. Thus, it is noted that although Murdoch does not share Christie's increase in indefinite nouns, they both show common lexical decline not found in James, validating the hypotheses with respect to lexical markers.

Although the analysis and data preparation was very carefully conducted, as the authors note, the data set is somewhat small with

[3] Distinctiveness Ratio is a measure of variability defined by the rate of occurrence of a word in a text divided by its rate of occurrence in another.

only 1–2 people for each of the two conditions, leaving it unclear what aspect of the results are reliable, as each of the examined women could potentially be unrepresentative of their group. In addition, general language shift or stylistic change could also have had an influence on the observed change.

While the work by Le *et al.* (2011) considered symptoms of pathological linguistic decline, the study by Pennebaker and Stone (2003) (hereafter also: P&S) focused on aspects of regular and expected linguistic ageing. In particular, they proposed four hypotheses about the effect of ageing on language. Firstly, they suggested that ageing was associated with a drop in negative affect words and a slight increase in positive affect words (hypothesis 1). Further, social words and first-person plural pronouns[4] were hypothesized to decrease relative to a person's decrease in social networks (hypothesis 2). If ageing was associated with a greater concern with the past relative to the future, linguistic shifts from future to past tense as well as a reduction in references to time altogether could be expected (hypothesis 3). Finally, older people were predicted to use fewer cognitively complex words (cognitive mechanisms and causal, insight, and exclusive words), whereas markers of verbal ability were not expected to show either monotonic increases or decreases (hypothesis 4). P&S investigated how the age of a person affected these linguistic categories, with respect to two very different data sets: one based on self-reports from emotional disclosure studies (the 'Disclosure project'; hereafter also: DP) and the other based on collected works of ten different authors across their individual life spans, hereafter also referred to as the 'Author project' (AP).

The Disclosure project featured 3,280 participants from 45 separate studies, of which 32 were traditional emotional disclosure experiments in which participants were randomly assigned to write about either a traumatic or emotional topic, or a superficial topic in the case of the controls (for details, see Pennebaker and Stone 2003). Although this data is ordered by age of participants, the samples may have originated from the same time period. Both DP and AP were assessed using correlation analysis and in addition the DP was also analyzed through simple linear and quadratic regression.

[4] These are hereafter also referred to as '1PL'.

Table 1: P&S's results showing means over individual age-variable correlations. Significance t-tests are based on means of the within-author (individual variable) correlations with age for the Author project and between-subject with age for the Disclosure project. Significance levels are indicated by: *: $p \leq 0.05$/**: $p \leq 0.01$/ ***: $p \leq 0.001$

LIWC variable	Example	AP	DP	
			Experimentals	Controls
Social and identity				
First-person singular	*I, me, my*	−0.26*	−0.18*	−0.18*
First-person plural	*we, us, our*	0.03	−0.01	−0.27*
Time orientation				
Past-tense verbs	*was, went, ate*	0.08	−0.20*	−0.22*
Present-tense verbs	*am, see, goes*	0.09	0.05	0.03
Future-tense verbs	*will, shall*	0.22*	0.19*	0.10*
Cognitive complexity				
Big words (> 6 letters)	*pontification*	0.10	0.35*	0.36*

Table 1 shows P&S's results for individual age-variable correlations for both data sets (limited to those variables that are also analyzed as part of the current research, as this paper only details a partial replication of the original study). For this, the results for the two DP conditions ('Experimentals'/'Controls') were based on between-subject analyses correlating each of the Linguistic Inquiry and Word Count (LIWC) variables with age. For the Author project, the correlation coefficient is based on mean within-author correlations between each author's age and the LIWC analyses for the works written at that age. As can be observed from the table, for first-person singular pronouns,[5] present and future tense, and big words, hereafter also referred to as long-letter sequences, all three correlations are in the same direction across the two sets, although only in two cases are all of them also significant (and the direction of first-person singular pronouns is inversely correlated with age while the other two variables directly correlate with age).

[5] These are hereafter also referred to as '1SG'.

Table 2 shows the collection of authors in the AP of the P&S study. Although it is balanced across genders, it contains some idiosyncrasies, such as the fact that most authors originated from Great Britain (England and Scotland), except for writers Louisa May Alcott and Edna St. Vincent Millay of American origin. Genre types include novels, plays and poetry, a fact that could present a confounding factor specifically for the analysis of pronouns that are usually distributed somewhat differently across these text types. The most relevant issue in this context is that authors' works are spread across five centuries (1591–1939) and language use would be expected to somewhat vary between the 16th and 20th centuries. It is to be assumed that this design was deliberate in order to extract very diverse samples – nevertheless, this may render them still less comparable and results could be spurious. In particular, if language has been affected by a continuous shift throughout this time, a significant effect in authors who did not compose language in parallel may still be attributable to general language change rather than ageing.

The final column in Table 2 shows the result of using regression weights for the LIWC variables based on the DP data to create an ageing coefficient for each individual author, which was then correlated with age. Thus, larger correlations signify more similarity to the DP analysis regarding the ageing variables. It is noticeable that five out of six significant correlations, i.e. Joanna Baille, Robert Graves, Edna St. Vincent Millay, William Wordsworth and William Butler Yeats, are based on genre types that could be more prone to irregularities, e.g. poetry and plays. Overall, neither analysis anchored in the DP or AP data is reported to have evaluated the influence of general language change.

Thus, apart from general language shifts, other possible confounding factors for the P&S study could have been introduced by the differences in pronoun distributions across varying text types as well as individual stylistic differences and developments, irrespective of any particular ageing process. In this work, we revisit the question of linguistic ageing for six variables previously analyzed. Specifically, we do not reanalyze P&S's data, but conduct a comparable experiment on a more temporally and genre-homogenous data set. We then compare our findings on the same variables to P&S's earlier results.

Table 2: P&S's 'Characteristics of Authors Chosen for the Author Project' (Pennebaker and Stone 2003, p. 297). F = female; M = male. The ageing coefficient correlations are within-subject simple correlations between each author's age and the ageing coefficient and were based on the regression weights from the Disclosure Project. Significance levels are indicated as follows: †: $p \leq 0.08$/*: $p \leq 0.05$/**: $p \leq 0.001$

Author	Nationality	Sex	Life span	Productive years	Genre	Analyzed works (n)	Words per work (M)	Ageing coefficient correlation
Louisa May Alcott	US	F	1832–1888	1854–1886	Novels, stories	19	40,273	−0.05
Jane Austen	England	F	1775–1817	1787–1817	Novels, stories	13	68,120	0.23
Joanna Baille	Scotland	F	1762–1851	1789–1827	Plays	20	18,921	0.60**
Charles Dickens	England	M	1812–1870	1836–1870	Novels	15	257,777	−0.23
George Eliot	England	F	1819–1880	1859–1876	Novels, stories	10	157,751	0.63*
Robert Graves	England	M	1895–1985	1910–1975	Poetry	100	1,689	0.18†
Edna St. Vincent Millay	US	F	1892–1950	1917–1947	Poetry	21	3,850	0.72**
William Shakespeare	England	M	1564–1616	1591–1613	Plays	37	22,975	0.03
William Wordsworth	England	M	1770–1850	1785–1847	Poetry	64	6,074	0.37**
William Butler Yeats	England	M	1865–1939	1889–1939	Poetry	34	2,217	0.40*

Note: For most novels, stories, and plays, each work was analyzed separately. For poetry, a work was defined by the various poems written within a given year. Exceptions include poems or collections that were known to have been written over several years, which were entered as separate text files.

DATA 3

The data analyzed for this research is divided into two main sets: twenty-two literary authors, comprising ten women and twelve men, and a corresponding reference corpus for the same time period. Table 3 shows the set of literary authors, all of whom published work between 1847–1923.[6] The corpus was populated in the following way: first

Table 3: Corpus of literary authors, indicating timeline, gender, number of works, size of works in megabytes and their total word count

Author	Timeline	Gender	Works	Size (MB)	Word count
Alice Brown	1884–1922	F	12	5.7	1064566
Amanda Minnie Douglas	1866–1914	F	51	24.5	4500421
Constance Fenimore Woolson	1873–1895	F	12	6.7	1204937
Edith Wharton	1897–1920	F	10	3.5	609351
Elizabeth Stuart Phelps Ward	1866–1907	F	21	5.8	1055611
Gertrude Atherton	1888–1923	F	19	9.1	1628163
Harriet Beecher Stowe	1852–1886	F	18	11.2	2049014
Louisa May Alcott	1854–1893	F	16	5.6	1027950
Marion Harland	1854–1914	F	15	9.0	1572983
Susan Warner	1850–1884	F	29	18.6	3467028
Charles Dudley Warner	1872–1899	M	14	6.1	1088452
Edgar Saltus	1884–1919	M	17	3.6	650825
Francis Marion Crawford	1882–1908	M	41	23.3	4238660
Harold McGrath	1903–1922	M	15	5.3	945365
Henry James	1877–1917	M	32	17.3	3123582
Horatio Alger	1866–1906	M	37	10.3	1840445
Mark Twain	1869–1916	M	23	11	1990085
Robert W. Chambers	1894–1922	M	38	20	3465933
Timothy Shay Arthur	1847–1890	M	30	10.7	1933432
Upton Sinclair	1898–1922	M	17	8.6	1572977
William Dean Howells	1867–1916	M	38	16.7	3063271
William Taylor Adams	1855–1896	M	49	17.5	3208971

[6] The corpus is motivated and described in more detail by Klaussner and Vogel (2018a). The data set is available at `http://www.scss.tcd.ie/clg/DCLSA/` – last verified October 2021.

the prolific authors Mark Twain and Henry James were chosen, which was inspired by several sources that suggested they may be interesting to contrast (Beach 1918; Canby 1951). The remaining contemporaneous authors were selected by first assembling a list of male and female American authors of the 19^{th}–20^{th} century using Wikipedia[7] and then selecting a subset of these authors, all of who had a few long works publicly available and spread out over at least twenty years. Also, for the purpose of estimating stable word distributions, shorter works of less than 150 kilobytes in length were excluded. In terms of temporal alignment, a fair subset of the authors wrote largely in parallel. For instance, Harriet Beecher Stowe, Louisa May Alcott, Marion Harland and Susan Warner all have their first work in this corpus within four years of each other (1850–1854).[8] Elizabeth Stuart Phelps Ward and Amanda Minnie Douglas both began writing about 15 years later in 1866.

The literary prose texts were mainly collected from Project Gutenberg (PG):[9] this part of the corpus consists of 397 hand-transcribed works; it was supplemented with 158 scanned works from the Internet Archive (IA).[10] In general, we might prefer to choose a hand-transcribed version of a text from Project Gutenberg rather than the possibly more noisy OCR version from the Internet Archive. However, in this case acquiring data with a time stamp close to the first publication date was essential and for this reason and especially when the equivalent PG version did not have a time stamp, the IA version was chosen instead if available. On occasion, the OCR versions were manually corrected, but this was determined on an individual basis and through human inspection only.

All data was prepared by manually removing parts that were written at a different time from the main work, along with introductions or

[7] `https://en.wikipedia.org/wiki/Category:19th-century_American_writers` – last verified October 2021.

[8] When using descriptions, such as *first* or *last* with respect to authors' works, this is generally to be understood with respect to this corpus; there might be cases where an earlier or later work for an author exists, but could not be included in this corpus.

[9] `https://www.gutenberg.org/` – last verified October 2021.

[10] `https://archive.org/` – last verified October 2021.

comments not by the author, such as copyright headers/footers, notes or introductions by editors. Additionally, tables of contents were also removed, as these do not usually follow a normal sentence structure. Klaussner and Vogel (2018a) provides more specific descriptions of the data and its basic pre-processing. The publication date of a text was set by taking the first documented date, e.g. first copyright or publication date, unless a preface clearly stated that the work had been subject to explicit revisions. The issue with dating in this case is that either dating a work too early or too late would distort the results.

The reference language corpus for the current work was assembled by taking an extract from The Corpus of Historical American English (COHA: Davies 2012).[11] COHA is a 475-million word corpus that contains samples of American English from 1810–2009, balanced in size, genre and sub-genre in each decade (1000–2500 files each). Depending on the particular type of analysis, different excerpts from the entire data set were used. The corpus contains balanced language samples from fiction, popular magazines, newspapers and non-fiction books, which are again balanced across sub-genre, such as drama and poetry.[12] While the corpus is balanced overall, some years contain proportionally more data from certain genres than others, where we observed strange frequency effects. However, to the best of our knowledge, for our current requirements of providing an approximation to general language usage at the time, this corpus still provided the best option.

METHODS 4

Section 4.1 describes how features were extracted, and is followed by Section 4.2: the statistical models used for the analysis.

[11] A free web-based version is accessible on: `https://www.english-corpora.org/coha/` – last verified October 2021.

[12] There is an Excel file with a detailed list of sources available on: `https://www.english-corpora.org/coha/` – last verified October 2021.

4.1 Feature extraction

We begin by describing the feature extraction adopted by Pennebaker and Stone (2003), interlaced with our own design, where modifications were deemed necessary. As previously mentioned, Pennebaker and Stone (2003) based their analysis on the LIWC system, whose categorization scheme is generally not openly accessible. This renders replication of less objective linguistic variables, such as negative or positive emotion words difficult.[13]

Table 1 only lists examples of non-reflexive uses of pronouns and main tenses, so it is unclear whether reflexive pronouns were included and how complex verb forms also indicating aspect, such as present perfect or future perfect, were treated in their analysis. For extracting 1SG/1PL pronouns in the current work the word was used in conjunction with the part-of-speech tag to identify the correct items, e.g. to avoid uses of *I* that refer to numbering.[14] As our experiments did not show differences between including or excluding reflexive pronouns, this analysis only reports on non-reflexive pronoun types.

Originally, P&S also included what they refer to as "time-related" words, such as *clock*, *hour* and *soon*. One can assume that they would also include temporal adverbs in general like *yesterday* or *today*. These temporal expressions may change the interpretation of regular tenses and could result in shifts between them. However, this may not be a trivial problem, as sometimes the overall tense would be more strongly signaled by the temporal adverb, e.g. examples (1) and (2), whereas in other cases the verb would be the determining factor, as in example (3).

(1) *She's there tomorrow.*

(2) *She's there today.*

(3) *She was there today.*

---

[13] To the best of our knowledge, these words were classified by several different students and can be (indirectly) accessed through the LIWC program. Research papers usually only provide examples rather than exhaustive lists.

[14] For all the computations in this work, the statistical programming language R (R. Core Team 2014) and associated packages were used. For POS-tagging the *NLP* (Hornik 2016) and *openNLP* (Hornik 2015) packages were used.

This suggests the need for a more intricate classification system than could be done justice as part of the present work. Here we resort to only using verb tenses to approximate the overall tenses. The main effect of not including temporal adverbs may be a shift from future to present tense counts. In order to approximate tense representation, we adopted the following classification: while POS tags could be used to directly identify some of the simpler tenses, this would not suffice to always correctly determine the difference between the present or present perfect tense usage of *have* and neither could it identify occurrences of the *going-to* future tense, as this is not marked explicitly on *going-to*.[15] To be able to make these distinctions, we used chunk tags to extract verb phrases and then analyzed the combination of tags within to determine the type of tense. In this, several sub-types corresponding to finer shades of difference in meaning are classified into the three main categories (*past/present/future*), as follows. The *present* type includes: simple present, present progressive, and conditional and modal variants, such as *can/could/may go*. The *past* type captures simple past, present perfect, past perfect, past progressive and, as with the *present* type, conditional and modal variants, such as *could have gone*. Finally, the *future* type covers simple future construction, such as *will/shall go* and *going to go*, but also *will have gone*. Finally, we define long-letter sequences as previously, as words whose length is greater than or equal to six letters.

After extracting the relevant features, texts in each corpus were combined by considering the year of publication, thereby reducing each set to one file per year per (author) corpus. Relative frequencies for each feature type were calculated by considering the ratio of the occurrence of the feature and all tokens for the same year. In addition, ordinal variables were created corresponding to year of publication (*year*), age of author at publication of text (*age*) and a categorical variable indicating the author (*A*) of a text.

[15] Still, somehow items such as *I'm going to school* had to be distinguished from *I'm going to go to school.*

4.2 Statistical modelling

This section describes aspects connected to the statistical analysis, i.e. regression models and standardization techniques, before moving on to model assessment.

Temporally-ordered data can be analyzed in different ways, for instance relating a variable to itself at different points in time as part of a "time-series" model or, as in the present case, by considering other variables at the same point in time thereby using an "explanatory model". Consequently, the prediction of a variable y is based on a function over a set of distinct variables: $x_1, x_2, \ldots, x_{p-1}, x_p = X$, with $y \notin X$, at the same time point $t : \{t \in 1, \ldots, n\}$, and some error term: $y_t = f(x_{1t}, x_{2t}, \ldots, x_{p-1t}, x_{pt}, error)$.

The regression models computed in the following experiments vary with respect to the data set used and whether individual author variation had to be accounted for. The reference corpus (RC) does not contain an *age* variable and is only evaluated with respect to *year* of publication, which serves to check whether a particular variable of interest is likely to have changed in relative frequency over time.

However, when analyzing the literary authors corpus, both *age* and *year* have to be considered as predictors, since the authors will align differently depending on the variable, i.e. James and Twain were not the same age in the same year. Thus, in order to argue for an ageing effect to be present for an individual, it has to (also) be found in a combined model of the authors, clearly outperforming the equivalent year-based model that does not depend on age, but may capture stylistic changes over time instead.

When analyzing different authors at the same time, one may have to resort to random effects models to account for individual variation between authors as shown by Equation (1), where y_{tj} is the response variable for author j at time t, x_{tj} is the individual-specific random effect and A_j is the author-specific random effect; ε_{tj} represents the error term. Similarly, Equation (2) shows the same for the quadratic model, adding predictor $\beta_2 x_{tj}^2$.

(1) $$y_{tj} = \beta_0 + \beta_1 x_{tj} + \beta_2 A_j + \varepsilon_{tj}$$

(2) $$y_{tj} = \beta_0 + \beta_1 x_{tj} + \beta_2 x_{tj}^2 + \beta_3 A_j + \varepsilon_{tj}$$

For fitting linear and normally distributed models, the *nmle* R package was used (Pinheiro *et al.* 2013). Data that was only log-normal was fitted through the *glmmPQL* function in the *MASS* package (Venables and Ripley 2002). In order to preserve similarity with P&S's study, the predictors *age* and *year* were standardized two-ways, one by computing z-scores, i.e. subtracting the mean and dividing by one standard deviation for the simple linear regression models, and also by taking the absolute value of the difference from the mean over the sample for the quadratic models. For correlation analysis, either Pearson correlation coefficient r or Spearman's ρ were used, for normally and non-normally distributed data, respectively.

The decision as to what type of model and correlation measure to use, i.e. parametric or non-parametric, was based on whether the linear model fulfilled all model assumptions: all models were tested for normality, kurtosis, skewness, nonlinear link function (for testing linearity) and heteroscedasticity.[16]

EXPERIMENTS 5

This section begins by examining background language change with respect to the six linguistic variables outlined in Table 1. Having considered background language change, Section 5.2 then investigates how these effects can be explicitly modelled in the case of the literary authors. This also allows us to determine to what extent background language may be responsible for effects observed in the individuals.

Background language change 5.1

Examining the change in linguistic variables over time raises the question to what extent these variables were subject to other outside in-

[16] Computed through the *gvlma* package in R (Pena and Slate 2014).

fluences, especially when considering a time span of ∼40 years or more. To be able to assign meaning to measures of linguistic ageing, a separate analysis of the change in the background language is conducted as part of this section. In general, observed individual effects could be either subsumed by language change or rendered more significant if they happen to be in the opposite direction. Thus, taking background language into account can both lessen and strengthen individual effects.

Table 4 shows correlation results for the reference corpus and both P&S's Disclosure project and Author project. The results for computing simple linear (β) and quadratic (β^2) models are displayed only for the reference corpus alongside the DP as the same model computations were not available for the AP. Our reference corpus shares characteristics with both of P&S's studies in that it covers a similar length of time as the DP (∼70 years) and years contain multiple individual samples rather than a strict within-subject design. However, it is more comparable to the AP design in that it is genuinely sampled from different time periods, whereas some of the DP's data representing different age groups could have originated from the same time period. For this reason, we aim for a general comparison or replication rather than remaining very close to the original study.

Language change effects can be observed with respect to at least three of the six variables, and this is specifically notable in the case of 1PL pronouns and past tense, where the effect is in the same direction as for the DP, and the case of long-letter sequences, where effects are in the opposite direction for both of P&S's studies.

5.1.1 Change in pronouns

Figure 1 depicts 1SG and 1PL pronouns in the RC over the time span from 1830–1919.[17] As can be observed, 1SG pronouns slightly increase in relative frequency over time. All model parameters in Table 4 show a positive but non-significant trend over time.

Both P&S's studies have significant, but negative associations for 1SG pronouns over time. 1PL pronouns experience a highly significant decrease in relative frequency over the reference corpus, and

[17] As there were some sampling irregularities in the reference corpus around 1923, the years after 1919 were excluded, resulting in 90 years of data.

Table 4: This table shows correlation analysis (r) and main model coefficients for simple linear (β) and quadratic models (β^2) for both P&S's Disclosure and Author project, and the current 18^{th}–19^{th} century reference corpus. Items marked with '!' signal that linearity assumptions were violated. By default Pearson's r is used, but is replaced by Spearman's ρ for departures from linearity; this is indicated by a superscript ρ. Significance levels are indicated as follows: *: $p \leq 0.05$/**: $p \leq 0.01$/***: $p \leq 0.001$

LIWC variable	P&S: Disclosure project			P&S: Author project	**Reference corpus**		
	r	β	β^2	r	r	β	β^2
Social and identity							
First-person singular	−0.13**	−0.14**	−0.2	−0.26*	0.18^{ρ}	0.0005!	0.000003!
First-person plural	−0.12**	−0.13**	0.19**	0.03	$-0.60^{\rho***}$	−0.0006!***	0.0000008!
Time orientation							
Past-tense verbs	0.04**	−0.16**	0.01	0.08	$0.7^{\rho***}$	0.004!***	−0.000002
Present-tense verbs	−0.02	0.04*	0.06**	0.09	0.16	0.0003!	0.000005**
Future-tense verbs	0.00	0.14**	−0.02	0.22*	0.04	0.00001!	0.0000002
Cognitive complexity							
Long-letter seq. (> 6 letters)	0.13**	0.26**	−0.03	0.10	$-0.51^{\rho***}$	−0.02!***	−0.000004

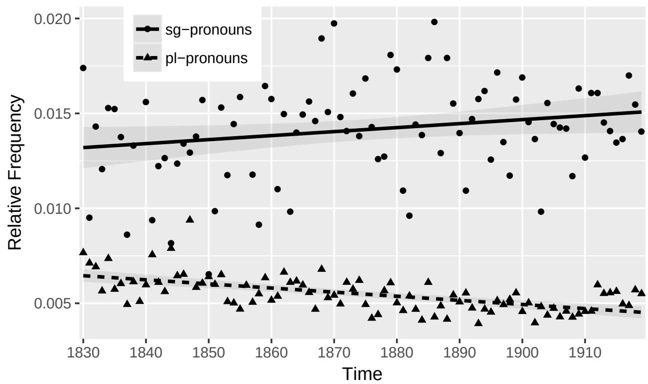


Figure 1: Reference corpus: 1SG and 1PL pronouns

while P&S's Author project reveals low non-significant correlations, their Disclosure project shares this highly significant downward trend. The linear model results mirror these correlations for both variables. There is less evidence of background language interference in the case of 1SG pronouns, but stronger indications in the case of 1PL pronouns.

5.1.2 Change in tenses

Figure 2 shows relative frequencies for past, present and future tense. Future tense shows little variation over time or at least not at a significant level, while examining Table 4 shows that both P&S's data sets have a positive association for future tense over time. Present tense appears stable in relative frequency and has a significant positive quadratic trend as can also be observed in P&S's DP. Past tense in

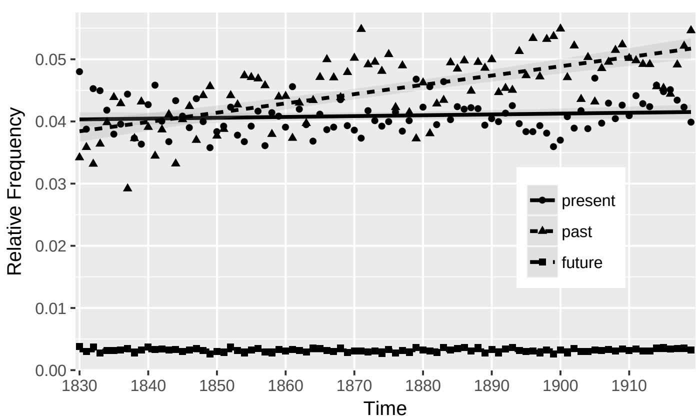


Figure 2: Relative frequencies of past, present and future tense

the RC has a highly significant positive correlation (0.7^{***}) and highly significant regression coefficient β, and while r is also positive and significant in P&S's DP, it is reported to have a significant negative linear regression coefficient (-0.16^{**}). Their AP has a non-significant positive correlation for both present and past tense. Both visual and statistical analysis indicate that the tenses, but especially the past tense, underwent change in frequency in background language use for the time period examined, and as with 1PL pronouns could therefore introduce noise into stylistic or ageing analyses.

Change in long-letter sequences 5.1.3

The development of long-letter sequences over the RC is shown in Figure 3. There is a continuous downward trend visible, which is confirmed by both a highly significant correlation coefficient ρ (-0.51^{***}) and a linear regression coefficient β ($-0.02^{!***}$) in Table 4. Both P&S's DP and AP have positive trends and therefore trends in the opposite direction (r of 0.13^{**} and 0.10 respectively).

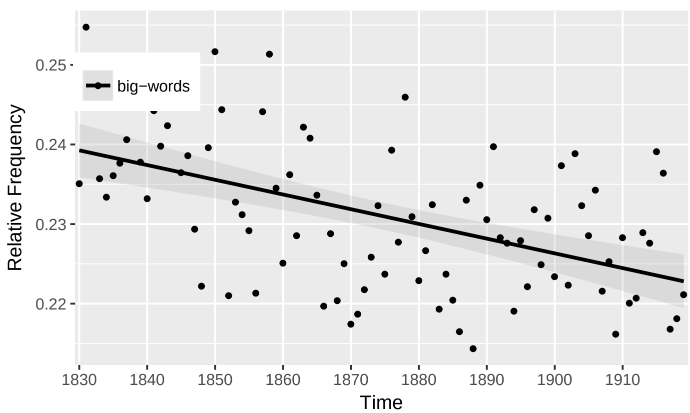


Figure 3: RC: long-letter sequences

Discussion 5.1.4

This section has examined six linguistic variables in a continuous section of general language usage that have been hypothesized in the literature to be affected by ageing in individual writers. Pennebaker and Stone (2003) found significant decreases in all their data sets with respect to 1SG pronouns. For the time frame examined here, no significant trend for 1SG pronouns based on publication year was observed in the reference language. This adds weight to the interpretation of the P&S 1SG effect as being one of ageing. 1PL pronouns were negatively associated with age for the Disclosure study and our reference corpus also showed a highly significant negative trend over time. This suggests that the 1PL effects may not be due to ageing. Pennebaker and Stone's work observed a significant decrease in past tense verbs in the DP, while this variable could be observed to increase in the RC. Present tense was not found to be a likely factor in ageing by P&S, which can be partially confirmed as the relative frequency did not seem to undergo a very pronounced shift. Similarly, there did not appear to be a very strong effect for future tense in our reference corpus, whereas it was found to increase over all of P&S's data sets, possibly implicating this as a real ageing effect. Long-letter sequences are comparable to the past tense situation: Pennebaker and Stone (2003) report a significant increase over their Disclosure project, whereas there is a significant decrease over the background language sample examined here. If their data were subject to similar effects, then this could render the linguistic ageing results more pronounced. This analysis has shown there to exist significant language change in most of the ageing variables examined. To what extent this challenges or amplifies results in the original study is not further examined here. Rather, the next section addresses how these underlying influences can be taken into account when examining linguistic ageing variables in the literary authors corpus, by attempting to estimate the impact of background language change more systematically for the literary authors. We then consider to what extent this underlying change influences interpretation of effects previously only attributed to ageing.

5.2 *Estimating impact of language change*

In this section, we aim to investigate the ageing hypotheses with respect to the literary authors corpus while controlling for background language influence. For instance, a random effects model as shown in Equation (3) can be used, taking into account reference language, where $\boldsymbol{ref}_{ij}$ is the relative frequency of the reference language for author j (A_j) at age i and random error ε_{ij}. Equation (4) shows the equivalent quadratic model.

(3) $$y_{ij} = \beta_0 + \beta_1 \boldsymbol{ref}_{ij} + \beta_2 Age_{ij} + \beta_3 A_j + \varepsilon_{ij}$$

(4) $$y_{ij} = \beta_0 + \beta_1 \boldsymbol{ref}_{ij} + \beta_2 Age_{ij} + \beta_3 Age^2_{ij} + \beta_4 A_j + \varepsilon_{ij}$$

The set of literary authors varied somewhat and for most variables only a subset of authors produced a normal or log-normal fit. For this reason different subsets of the entire data were used to test individual variables' hypotheses.

Table 5 shows the results of computing simple linear random effects models for the six linguistic variables. The first two columns show model coefficients for the age and background language predictors. The third column specifies what model type was used, i.e. normal (N) or log-normal (LN) and the final column lists the respective size of author set. Overall, there is little evidence for either a very strong

Table 5: This table shows the main model coefficients for simple linear regression using random effects models. 'Age.std' and 'Ref.std' refer to standardized age predictor and background change factor respectively. 'Model type' specifies normal (N) or log-normal (LN) setting and '|Authors|' refers to the size of the supporting set. Significance is indicated by: *: $p \leq 0.05$/**: $p \leq 0.01$/***: $p \leq 0.001$/ $^{\cdots}$: $p \leq 0.1$. A '†' on the ageing coefficient indicates that the equivalent model using *year* (of publication) was more significant

LIWC variable	Model coeff.		Model type	|Authors|
	Age.std	Ref.std		
Social and identity				
First-person singular	0.0008	−0.0002	N	12
First-person plural	0.02	$0.07^{\cdots}$	LN	15
Time orientation				
Past-tense verbs	0.0008	−0.0002	N	10
Present-tense verbs	0.00009	−0.0004	N	18
Future-tense verbs	$-0.0002^{***\dagger}$	0.0005	N	20
Cognitive complexity				
Long-letter seq. (> 6 letters)	0.0007	−0.002	LN	21

Figure 4: R output for a *glmmPQL*-based model predicting future tense from reference language and *age* or *year*

```
Fixed effects: response ~ Ref.std + Age.std
                    Value    Std.Error  DF   t-value p-value
(Intercept)  0.0029418574 2.019705e-04 335 14.565781  0.0000
Ref.std      0.0000592255 5.052461e-05 335  1.172211  0.2419
Age.std     -0.0002297820 5.675550e-05 335 -4.048629  0.0001

Fixed effects: response ~ Ref.std + Year.std
                    Value    Std.Error  DF   t-value p-value
(Intercept)  0.0029857001 0.0001912692 335 15.609939  0.0000
Ref.std      0.0000626322 0.0000505549 335  1.238895  0.2163
Year.std    -0.0003035166 0.0000706138 335 -4.298262  0.0000
```

influence of background language change or linguistic ageing. The only nearly significant reference language coefficient is 1PL pronouns. Figure 4 presents evidence for some language change influence, i.e. removing the reference language predictor causes the *Year.std* predictor to become significant, while the ageing predictor *Age.std* in the equivalent model does not become more important, indicating that time of publication remains more salient than age of author. The only significant ageing predictor is for future tense; however, considering the equivalent model using *year* (of publication) instead of *age* (at time of publication) renders an even more significant model, calling into question the validity of *age* as a main cause of the observed effect.

Table 6 shows the results for computing quadratic random effect models for the six variables based on Equation (4). Similarly to the simple linear model results, quadratic models also do not yield well fitting models (in terms of significant predictors) for either age or background language predictors. For 1PL pronouns, the reference language predictor is almost significant in the sense of very nearly crossing the threshold for statistical significance at the 95% confidence level, as in the case of the simple linear model in Table 5. Although the ageing predictor for future tense in Table 6 is not significant, the equivalent quadratic year predictor is.

Finally, we turn to the last part of this analysis, namely the question of stylistic differences between authors. For instance, one could consider the question of whether there is likely to be anything particular about Mark Twain's and Henry James' style development compared to the other authors given that these two have received considerable attention from literary scholars. Further, we consider the

Table 6: This table shows the main model coefficients for quadratic regression using random effects models. 'Age.std^2' and 'Ref.std' refer to standardized age predictor and background change factor respectively. 'Model type' specifies normal (N) or log-normal (LN) setting and '|Authors|' refers to the size of the supporting set. Significance is indicated by: *: $p \leq 0.05$ /**: $p \leq 0.01$/***: $p \leq 0.001$/ $^{\cdots}$: $p \leq 0.1$. A '†' on the ageing coefficient indicates that the equivalent model using *year* (of publication) was more significant

LIWC variable	Model coeff.		Model type	|Authors|
	Age.std^2	Ref.std		
Social and identity				
First-person singular	−0.00001	−0.5	N	13
First-person plural	0.03	0.07$^{\cdots}$	LN	15
Time orientation				
Past-tense verbs	−0.0002	0.1	N	10
Present-tense verbs	−0.00001	−0.02	N	18
Future-tense verbs	−0.000001†	0.2	LN	18
Cognitive complexity				
Long-letter seq. (> 6 letters)	0.005	−0.002	LN	21

specific case of first-person pronouns. Figure 5, Figure 6, Figure 7 and Figure 8 show 1SG and 1PL pronouns for Twain and James alongside some of the other authors in the set, as well as a line representing the average over all authors in the set.[18] Figure 5 shows James and William Dean Howells and Figure 6 shows Twain and Elizabeth Stuart Phelps Ward. For neither Twain nor James does there appear to be a particular development in the form of a trend for 1SG pronouns. Nor is their level of variation around the authors' average among the highest. As the plots indicate, Howells and Ward show more variation for 1SG pronouns than either Twain or James. Figure 7 and Figure 8 confirm this general impression. For 1PL pronouns, James shows comparatively little variation over time, while Twain's style displays

[18] The 'aut-ref' line represents an average over all authors in the set, computed by, for each year, taking the raw frequencies for that year and two years before and after for each author separately, then averaging over all tokens in those years. Given this set of relative frequencies for a feature, the final frequency is given by averaging over all authors for a given year. Hereafter, this is also referred to as 'author reference corpus' or 'ARC'.

Figure 5: 1SG pronouns for James, Howells and the ARC

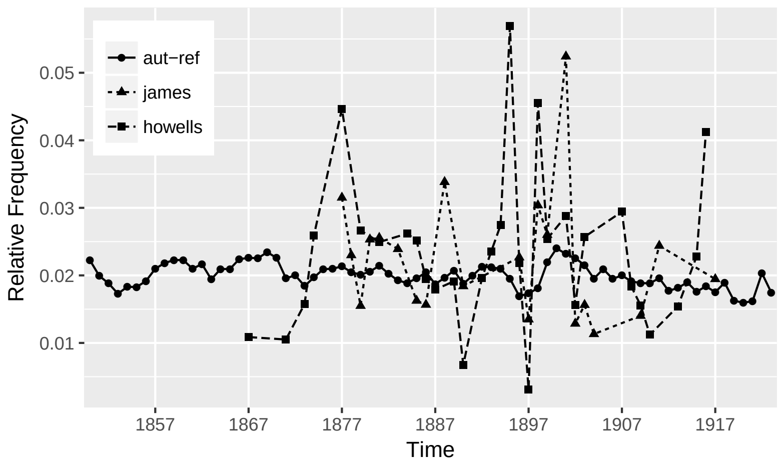
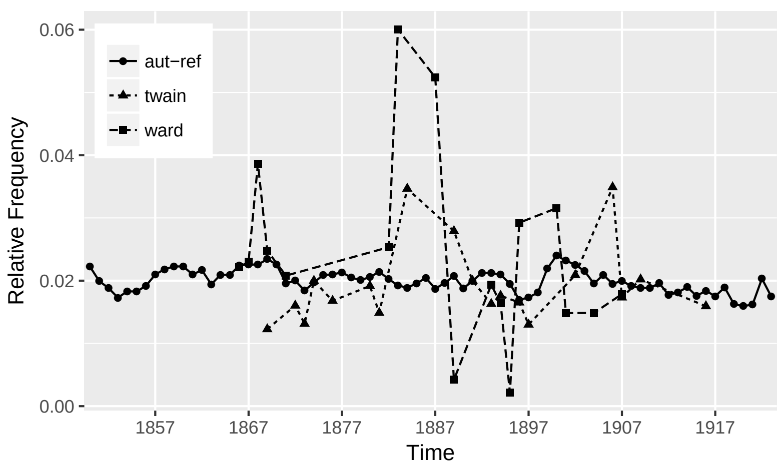


Figure 6: 1SG pronouns for Twain, Ward and the ARC



somewhat more variation around the authors' average. However, both authors are not unique in their tendencies. Like James, Alice Brown deviates comparatively little from the average, while Timothy Shay Arthur's relative frequency also increases in his last works similarly to Twain. Thus, there appears to be little evidence that Twain and James are decidedly different from their contemporaries in terms of style change. In the previous section, we identified an effect for 1PL with respect to background language effect, yet overall there is little evidence that there is a systematic influence of age or background language for these literary authors, at least for the variables examined.

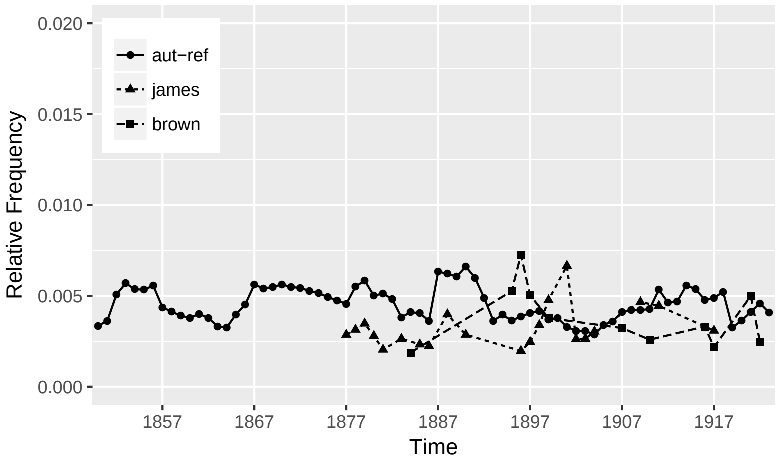


Figure 7:
1PL pronouns for James, Brown and the ARC

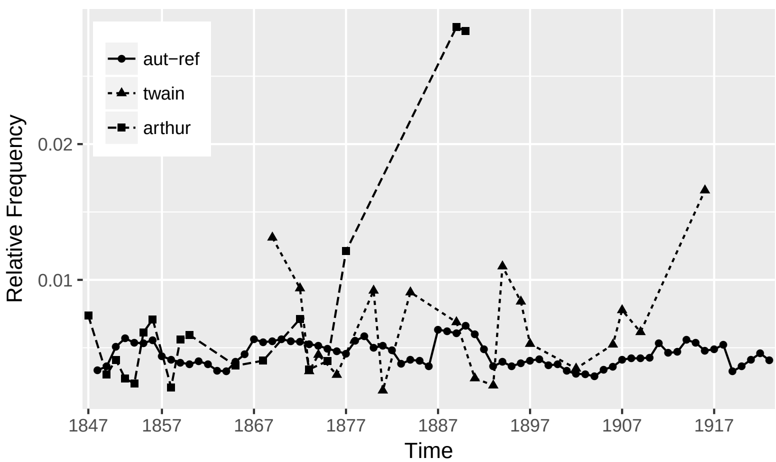


Figure 8:
1PL pronouns for Twain, Arthur and the ARC

This could indicate that literary authors have a higher command over their language usage and may be more impervious to outside influences.

DISCUSSION 6

This work has considered aspects of linguistic ageing and how this influences literary authors. In part, the study presented here was a replication of an earlier study by Pennebaker and Stone investigating the ageing effects in emotional disclosure studies and a corpus of literary authors. Although significant effects were found with respect to pronouns, future and past tense, and long-letter sequences in their study, these results did not replicate with respect to the authors examined here in a unified fashion that would suggest a rise or fall in frequency is actually due to age rather than only stylistic variation of individual authors. The fact that the results of the earlier study could not be replicated may be due to properties of this particular data set, but it could also hint at the possibility of this linguistic ageing effect not existing for professional writers, who could conceivably possess a higher command over their language style than non-professional writers. This would be consistent with P&S's findings insofar as their results for literary authors were also less significant than those for non-professional writers. This does not necessarily challenge the existence of linguistic ageing as a phenomenon, but rather suggests that the variables analyzed here do not provide good proxy measures for it, at least not with respect to literary writers. However, for this analysis no other non-linear models have been examined, something that would have to be done to completely refute the proposed hypotheses with respect to ageing.

The other purpose of this study was to examine these six variables for evidence of language change, and the results indicate significant change in the usage of at least 1PL pronouns, past and present tense verbs, and long-letter sequences. Overall, the models computed above for the literary authors present little evidence that background language (change) had a strong influence on them. However, the models built for 1PL pronouns present some evidence of background language influence, which indicates the necessity to control for it in general. A final result of this analysis was the diversity in the literary authors, which interestingly was not (only) caused by the prominent writers Mark Twain and Henry James. Instead, our analysis suggests that overall they seemed to align well with their contemporaries.

Based on this analysis, it appears that there could be some variation between authors for the six variables examined, possibly indicating stylistic differences with respect to other variables. These differences could be explored in more depth by looking more generally at stylistic change in the literary authors against the backdrop of general language shifts.

CONCLUSION 7

This work has considered to what extent ageing affects language development, examining six linguistic variables that had been reported as significant in the literature. While effects in previous studies were mainly found for non-professional writers, even significant effects confirmed by P&S for literary authors could not be replicated here. This does not necessarily prove an absence of previously identified effects, but calls for additional research to investigate this further. There is strong evidence of background language change for these variables, calling for explicit modelling of this influence, as has been exemplified as part of this work.

ACKNOWLEDGEMENTS

This research is supported by Science Foundation Ireland (SFI) through the CNGL Programme (Grants 12/CE/I2267 and 13/RC/2106) in the ADAPT Centre (`https://www.adaptcentre.ie`).

REFERENCES

Joseph Warren BEACH (1918), *The method of Henry James*, Yale University Press.

Helen BIRD, Matthew A. LAMBON RALPH, Karalyn PATTERSON, and John R. HODGES (2000), The rise and all of frequency and imageability: noun and verb production in semantic dementia, *Brain and Language*, 73(1):17–49, `https://www.sciencedirect.com/science/article/pii/S0093934X00922934`.

Deborah M. BURKE and Meredith A. SHAFTO (2008), Language and ageing, *Science*, 346:373–443.

Henry Seidel CANBY (1951), *Turn west, turn east: Mark Twain and Henry James*, Biblo & Tannen Publishers.

Mark DAVIES (2012), The 400 million word Corpus of Historical American English (1810–2009), in *English Historical Linguistics 2010: Selected Papers from the Sixteenth International Conference on English Historical linguistics (ICeHl 16), Pécs, 23–27 August 2010*, pp. 231–61.

Richard S. FORSYTH (1999), Stylochronometry with substrings, or: a poet young and old, *Literary and Linguistic Computing*, 14(4):467–478.

David L. HOOVER (2007), Corpus stylistics, stylometry, and the styles of Henry James, *Style*, 41(2):174–203.

Kurt HORNIK (2015), *openNLP: Apache OpenNLP tools interface*, `https://CRAN.R-project.org/package=openNLP`, R package version 0.2-5.

Kurt HORNIK (2016), *NLP: natural language processing infrastructure*, `https://CRAN.R-project.org/package=NLP`, R package version 0.1-9.

Susan KEMPER, Lydia H. GREINER, Janet G. MARQUIS, Katherine PRENOVOST, and Tracy L. MITZNER (2001), Language decline across the life span: findings from the nun study, *Psychology and Aging*, 16(2):227–239.

Carmen KLAUSSNER and Carl VOGEL (2018a), A diachronic corpus for literary style analysis, in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, European Language Resources Association (ELRA).

Carmen KLAUSSNER and Carl VOGEL (2018b), Temporal predictive regression models for linguistic style analysis, *Journal of Language Modelling*, 6(1):175–222.

Xuan LE, Ian LANCASHIRE, Graeme HIRST, and Regina JOKEL (2011), Longitudinal detection of dementia through lexical and syntactic changes in writing: a case study of three British novelists, *Literary and Linguistic Computing*, 26(4):435–461.

Erez LIEBERMAN, Jean-Baptiste MICHEL, Joe JACKSON, Tina TANG, and Martin A. NOWAK (2007), Quantifying the evolutionary dynamics of language, *Nature*, 449(7163):713.

Edsel A. PENA and Elizabeth H. SLATE (2014), *gvlma: global validation of linear models assumptions*, `http://CRAN.R-project.org/package=gvlma`, R package version 1.0.0.2 (last verified: 24.08.2015).

James W. PENNEBAKER and Lori D. STONE (2003), Words of wisdom: language use over the life span, *Journal of Personality and Social Psychology*, 85(2):291–231.

Jose PINHEIRO, Douglas BATES, Saikat DEBROY, Deepayan SARKAR, and R. CORE TEAM (2013), *nlme: linear and nonlinear mixed effects models*, R package version 3.1-113.

R. CORE TEAM (2014), *R: a language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, `http://www.r-project.org`.

Sanja ŠTAJNER and Ruslan MITKOV (2011), Diachronic stylistic changes in British and American varieties of 20th-century written English language, in *Proceedings of the Workshop on Language Technologies for Digital Humanities and*

Cultural Heritage at RANLP, pp. 78–85,
https://www.aclweb.org/anthology/W11-4112/.

Constantina Stamou (2007), Stylochronometry: stylistic development, sequence of composition, and relative dating, *Literary and Linguistic Computing*, 23(2):181–199.

W. N. Venables and B. D. Ripley (2002), *Modern applied statistics with S*, Springer, fourth edition, http://www.stats.ox.ac.uk/pub/MASS4.

Carmen Klaussner
0000-0002-5469-2167
klaussnc@tcd.ie

Carl Vogel
vogel@tcd.ie

Arnab Bhattacharya
0000-0003-2828-7691
bhattaca@tcd.ie

School of Computer Science
and Statistics, O'Reilly Institute,
Trinity College Dublin, Ireland

Carmen Klaussner, Carl Vogel, and Arnab Bhattacharya (2021), *Aspects of linguistic ageing in literary authors across time*, Journal of Language Modelling, 9(2):195–223
https://dx.doi.org/10.15398/jlm.v9i2.270

This work is licensed under the *Creative Commons Attribution 4.0 Public License*.
http://creativecommons.org/licenses/by/4.0/

Extraposed relative clauses in Role and Reference Grammar. An analysis using Tree Wrapping Grammars

Laura Kallmeyer
Heinrich-Heine-Universität Düsseldorf

ABSTRACT

This paper proposes an analysis of extraposed relative clauses in the framework of Role and Reference Grammar (RRG), adopting its formalization as a tree rewriting grammar, specifically as a Tree Wrapping Grammar (TWG). Extraposed relative clauses are a puzzle since the link to the antecedent noun can be rather non-local but it seems nevertheless appropriate to model it as a syntactic dependency and not a purely anaphoric relation. Moreover, certain types of determiners require their NP to be modified by a (possibly extraposed) relative clause, and any comprehensive framework should account for this. We show that the tree wrapping operation of TWG, which is conventionally used to fill argument slots out of which some elements have been extracted, can be used to model extraposed relative clauses. The analysis accounts for the non-locality of the phenomenon while capturing the link to the antecedent NP in a local way (i.e., within a single elementary tree).

Keywords: Role and Reference Grammar, tree rewriting grammars, extraposed relative clauses, extended domain of locality

INTRODUCTION 1

This paper makes two contributions: first, it proposes a precise and well-defined analysis of extraposed relative clauses within the grammar theory of *Role and Reference Grammar* (RRG; Van Valin and LaPolla 1997; Van Valin 2005); and, second, by doing so it develops an analysis of this phenomenon within a tree rewriting grammar formalism in the spirit of *Lexicalized Tree Adjoining Grammar* (LTAG Joshi and Schabes 1997; Abeillé and Rambow 2000) while overcoming the limitations of LTAG when dealing with extraposition.

Extraposed relative clauses are a challenge for any grammar theory due to the possible non-locality of the link between the relative clause and the antecedent (see Walker 2017, for an extensive description of the phenomenon). Some German examples of extraposed restrictive relative clauses are given in (1) (our own examples).[1]

(1) a. Es fängt der *Spieler* an, [*der zuletzt in Portugal*
EXPL starts the player PTCL, who most.recently in Portugal
war].
was
'The player is starting who was in Portugal most recently.'

b. Ich fahre mit dem *Freund* nach Portugal, [*der gestern*
I go with the friend to Portugal, who yesterday
das Spiel gewonnen hat].
the game won has
'I go to Portugal with the friend who won the game yesterday.'

c. Es fängt das Team des *Spielers* an, [*der zuletzt*
EXPL starts the team the player PTCL, who most.recently
in Portugal war].
in Portugal was
'The team of the player is starting who was in Portugal most recently.'

[1] Throughout the paper, antecedent noun and relative clause are both in italics, and the relative clause is in additional brackets. In sentences with more than one relative clause, additional indices indicate the respective antecedent-modifier relations.

In some places, abbreviations are used that follow the Leipzig Glossing Rules (Lehmann 1982), for instance EXPL for 'expletive'. Less standard, PTCL in (1) stands for 'particle'.

d. Es fängt die Figur aus dem Team desjenigen *Spielers*
EXPL starts the figure from the team the.one player
an, [*der zuletzt in Portugal war*].
PTCL, who most.recently in Portugal was
'The figure from the team of the player is starting who was in Portugal most recently.'

The antecedent in (1a) is an argument of the main verb, in (1b) it is part of an adjunct PP, and in (1c) and (1d) it is embedded in an argument. These examples illustrate that the antecedent of the extraposed relative clause is not necessarily an argument or modifier of the verbal head of the clause to which the relative clause attaches. It can be further embedded, and in principle there is no limit to the level of embedding (see (1d)). Consequently, one needs to find some "non-local" way for the antecedent NP and the relative clause to communicate with each other.

A further example of an embedded antecedent and an extraposed (non-restrictive) relative clause is the following, from Müller (2004), who also points out the non-local character of such dependencies.

(2) Karl hat mir [eine Kopie [einer Fälschung [des
Karl has me [a copy.ACC [a forgery.GEN [the
Bildes [*einer Frau*]]]] gegeben, [*die schon*
painting.GEN [a woman.GEN]]]] given, [who already
lange tot ist].
long.time dead is]
'Karl gave the copy of a forgery of a painting of a woman to me, who has been dead for a long time.' (Müller 2004)

Grammar theories that are able to establish non-local syntactic dependencies by percolating an arbitrary number of objects (for example a list of identifiers of antecedent NPs that might be modified by an extraposed relative clause) through the constituent tree can deal with such data. The main task is then to constrain the mechanisms for these non-local dependencies in appropriate ways (see Kiss 2005, Crysmann 2013 and Walker 2017 for an HPSG analysis along these lines). In contrast to this, grammar theories that assume an extended domain of syntactic locality, i.e., that have a set of elementary syntactic building blocks that each comprises a predicate together with its argument slots and adjunction sites for possible modifiers, would preferably choose a local analysis. In other words, they would group the antecedent NP and the extraposed relative clause (or its attachment site) into the same elementary unit. Such approaches, however, usually come with a formalization that assumes rather constrained composition operations for elementary structures, which results in restrictions concerning the non-locality of these dependencies. They therefore often have difficulties with the largely unrestricted character of extraposition. An example of such a formalism is LTAG (Joshi and Schabes 1997; Abeillé and Rambow 2000). To our knowledge, an analysis of extraposed relative clauses in LTAG has not yet been proposed. We discuss different options in Section 5.1 and show that, due to the restricted nature of LTAG's adjunction operation, the formalism is not able to account for extraposed relative clauses with an analysis that models the dependency between antecedent and relative clause as part of an elementary tree and that is in line with standard LTAG assumptions concerning grammar theory, i.e., concerning the form of elementary trees.

In this paper, we start from RRG (Van Valin and LaPolla 1997; Van Valin 2005), a grammar theory that has been shown to be adequate for describing a large range of typologically different languages. We adopt its formalization as a *Tree Wrapping Grammar (TWG)* (Kallmeyer *et al.* 2013; Kallmeyer 2016; Kallmeyer and Osswald 2017; Osswald and Kallmeyer 2018), a tree rewriting grammar along the lines of LTAG but with a larger generative capacity. We will show that this grammar formalism can model the relation between antecedent NP and relative clause as a local dependency, due to the expressive power of the tree wrapping operation.

Note that, in this paper, we only model the syntax of extraposed relative clauses; semantics is left aside. The main goal of the paper is, starting from RRG's assumptions about the form a constituent tree should have, to explain how this tree comes about. In other words, to develop a decomposition of the constituent tree into its elementary building blocks that captures all dependencies and constraints we want to model.

The remainder of this paper is organized as follows. The next section introduces RRG, then gives a more detailed overview of the data we are concerned with, and also introduces TWG and explains the way RRG is formalized. The analysis we propose for extraposed relative clauses is developed in Section 3, and Section 4 discusses different possibilities to model obligatory extraposed relative clauses. Section 5 compares our approach to others; and Section 6 concludes the paper.

PRELIMINARIES 2

Role and Reference Grammar 2.1

RRG is a non-transformational linguistic theory whose development has been strongly inspired by typological concerns and in which semantics and pragmatics play significant roles. The assumptions RRG makes concerning syntactic structure are guided by the question of what a linguistic theory would "look like if it were based on the analysis of languages with diverse structures such as Lakhota, Tagalog and Dyirbal [...]" (Van Valin 2005, page 1). That is, the syntactic structures underlying RRG cover among others free word order languages such as Dyirbal where a verb and its arguments and adjuncts can apparently be in any order (see Van Valin 2005, page 5 for an example) and where, therefore, a distinction between sentence or clause on the one hand and VP on the other hand does not seem appropriate. In general, RRG's syntactic structures are rather flat due to the aim to develop something applicable to all varieties of languages.

RRG's syntactic theory reflects semantic distinctions: One of the basic assumptions of RRG is that clauses have a *layered structure* which reflects the distinction between predicates, arguments, and non-arguments. The *core* layer (category CORE) consists of the *nucleus* (category NUC), which specifies the verb or rather the predicate, and its arguments. The *clause* layer (category CLAUSE) contains the core as well as extracted arguments. Each of the layers can have a *periphery* for attaching adjuncts. Furthermore, operators (e.g., temporal operators, definiteness operators, modals, etc.) are taken to be part of a separate operator projection which is nonetheless linked to the constituent structure. Each operator scopes over a specific layer. Other projections of predicative elements (NPs, APs, etc.) also come with nucleus and core layers. For such a category XP, the different layers are called NUC_X, $CORE_X$ and XP while for the entire clause, they are NUC, CORE, CLAUSE, and SENTENCE. The latter layer is added to clauses that have illocutionary force.

There are two treebanks of RRG structures currently under construction, which we use as sources for sample RRG trees: RRGbank (Bladier *et al.* 2018),[2] which constitutes an RRG-based annotation of parts of the Penn Treebank (PTB, Marcus *et al.* 1994), and RRGparbank (Bladier *et al.* 2020a),[3] a parallel treebank of Orwell's 1984 novel, based on the Multext-East 1984 corpus (MULTEXT-East "1984" annotated corpus 4.0, Erjavec *et al.* 2010), and extended with German and French. In the latter, besides English, there are also German, Russian, French and (to a lesser degree) Hungarian and Farsi RRG annotations. In these treebanks, operators and periphery elements are marked as such (category OP or category extension *-peri*) and they attach to the element they scope over/to whose periphery they belong. An example, taken from RRGbank,[4] is given in Figure 1 with two operators, a tense operator that attaches at the CLAUSE node, and a definiteness operator that attaches at the NP level, and two periphery elements, namely an adjectival modifier attaching at the corresponding $CORE_N$ and a modifier NP attaching at the CORE node.[5] Punctuation is omitted in the figure. This example is special in that it is a copula construction, therefore the nucleus ("be payable") is not a verbal predicate but a predication consisting of an auxiliary and an adjectival phrase.

The two treebanks mark extraposed relative clauses by a coreferential index REF = 1, REF = 2, etc., that is shared by the antecedent NP and the relative pronoun, which facilitates the search for these constructions. Many of the examples used in this paper are taken from these treebanks.

Concerning relative clauses, which are modifiers and, consequently, peripheral elements in RRG, RRG makes the following assumptions with respect to their categories and attachment sites: Depending on whether a relative clause is restrictive or not, it modi-

[2] `https://rrgbank.phil.hhu.de`
[3] `https://rrgparbank.phil.hhu.de`
[4] RRGbank sentence no. 3921, 12 Feb 2021.
[5] NP = nominal phrase, AP = adjectival phrase, QP = quantifier phrase.

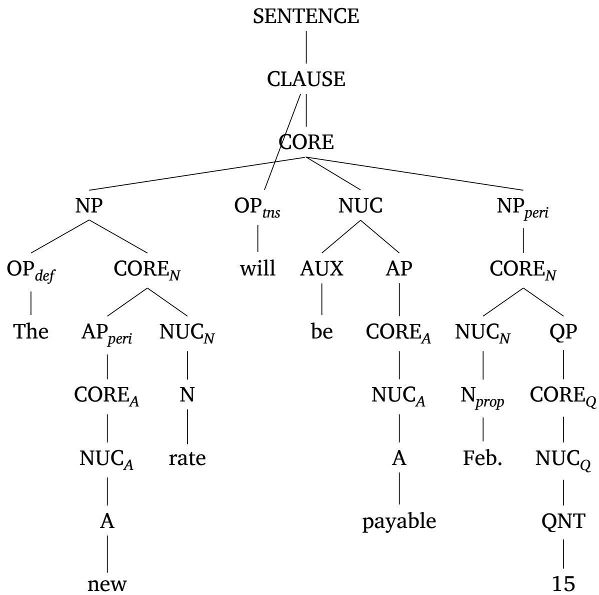


Figure 1:
Layered structure
of the clause, operators
and periphery elements

fies different parts of the NP. A restrictive relative clause provides an additional restriction on the predicate expressed by the NP's noun. Therefore, in RRG, restrictive relative clauses are considered to be part of the periphery of the nucleus of the NP (Van Valin 2005, Table 7.8, page 267). In contrast to this, non-restrictive relative clauses provide additional information about the NP's referent, therefore RRG considers them as being part of the periphery of the NP node (Van Valin 2005, Figure 6.29, page 222). Furthermore, a non-restrictive relative clause can have its own illocutionary force and is therefore treated as a SENTENCE constituent in RRG, while restrictive relative clauses are of category CLAUSE. Example (3) gives examples for both such types from the RRG treebanks. The corresponding RRG trees can be found in Figures 2 and 3 (punctuation is omitted).[6]

(3) a. "That's the *detail* [*that appeals to me*]."
(restrictive relative clause from RRGparbank)[7]

[6] PrCS = pre-core slot, a position mainly for extracted arguments.
[7] RRGparbank sentence no. 853, en, 12 Feb 2021.

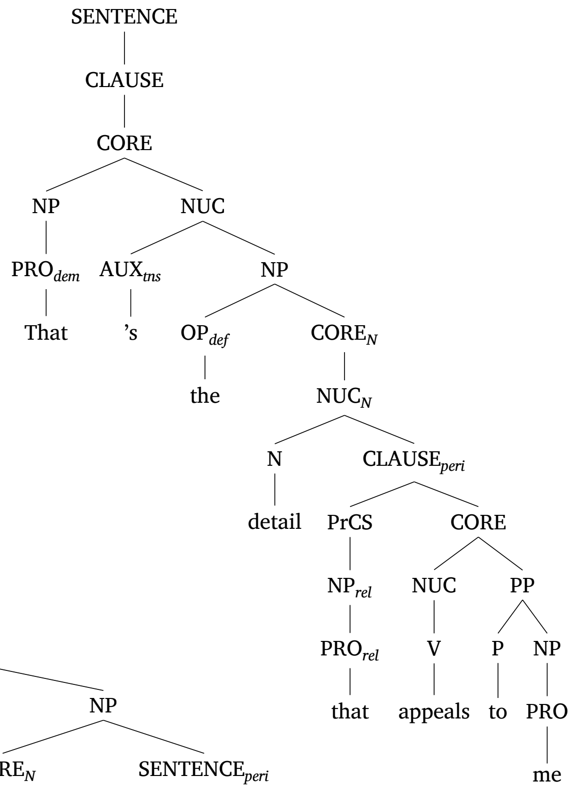


Figure 2: A restrictive relative clause from RRGparbank

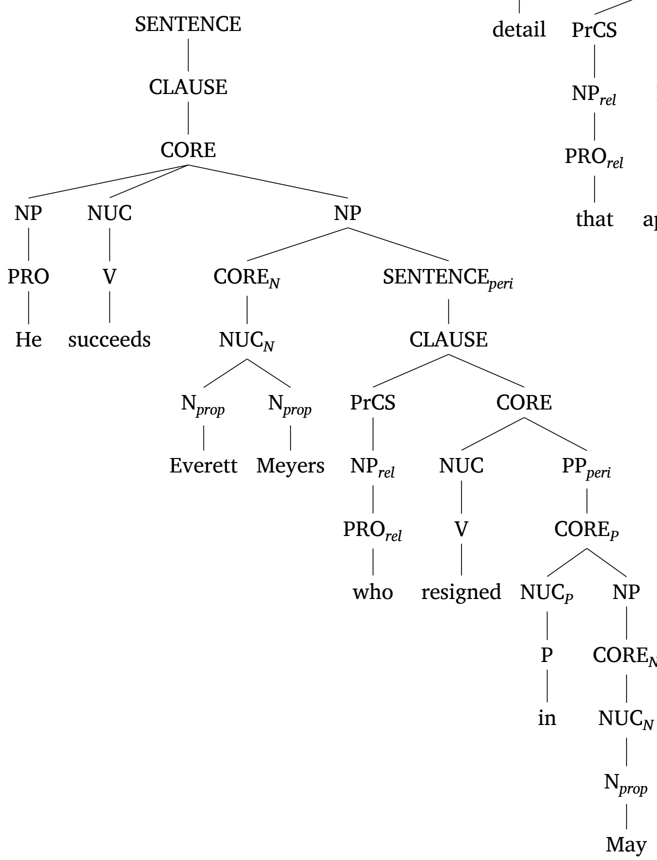


Figure 3: A non-restrictive relative clause from the RRGbank

b. He succeeds *Everett Meyers*, [*who resigned in May*]. (non-restrictive rel. clause, RRGbank)[8]

The focus of this paper is on restrictive relative clauses that have an overt antecedent NP, i.e., an NP that they modify.

Extraposed relative clauses: data 2.2

As mentioned above, restrictive relative clauses can not only appear inside the NP whose nucleus they modify but they can also be extraposed. Examples are (4b) (from Walker 2017) and (5) (from the RRG treebanks).

(4) a. *A girl* [*who was singing a song*] came in.
   b. *A girl* came in [*who was singing a song*].
   (Walker 2017, example (1), page 1)

(5) a. "You 've got *some minds* here [*that wo n't think progressively*]," he says.[9]
   b. Stratus Computer, which reported *earnings* late Friday [*that were in line with a disappointing forecast*], eased 3/4 to 24 on 816,000 shares.[10]
   c. "*Nothing* has happened [*that you did not foresee*]."[11]

In the RRG trees for the sentences in (5), the extraposed relative clause always attaches to the CLAUSE node that dominates the antecedent NP. The RRG tree for (5c) is given in Figure 4.

In (4b), we have the nucleus (*came in*) in between the NP and its relative clause, and the same holds for (5c). In (5a) and (5b) the constituents that separate the antecedent NP from its relative clause are modifiers of the predication (i.e., the CORE), namely the adverb *here* and the NP *late Friday* respectively.

[8] RRGbank sentence no. 40620, 12 Feb 2021.

[9] RRGbank sentence no. 31153, 12 Feb 2021. The separations of *You've* into two tokens *You* and *'ve* and of *won't* into *wo* and *n't* has been a choice of the tokenization of the treebank.

[10] RRGbank sentence no. 24269, 12 Feb 2021.

[11] RRGparbank sentence no. 5922, en, 12 Feb 2021.

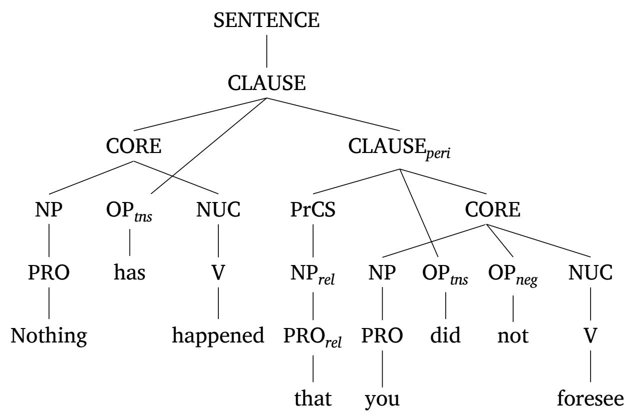


Figure 4: Tree from RRGparbank for an extraposed restrictive relative clause (punctuation omitted)

In (4) and (5), the antecedent NP is always an argument of the main verb, i.e., the link between antecedent NP and extraposed relative clause is still local in the sense that the attachment sites for antecedent NP and relative clause are part of the same layered structure, which means that there is a single NUC–CORE–CLAUSE spine such that the antecedent NP is an argument node immediately below CORE and the relative clause attaches at the CLAUSE node. However, as pointed out, among others, by Kiss (2005), Crysmann (2013), Holler (2013) and Walker (2017), this is not always the case: we can also have extraposition in cases where the antecedent NP is embedded within a PP below CORE while the relative clause attaches at the CLAUSE node, both with peripheral (i.e., modifying) PPs as well as with argument PPs, as in (6).

(6) a. I saw it [in a *magazine*]$_{\text{PP}_{\text{peri}}}$ yesterday [*which was lying on the table*].
(Baltin 1978, example (138), page 115)
b. I arrived [at a *solution*]$_{\text{PP}_{\text{arg}}}$ yesterday [*which I found totally unsatisfying*].
(Baltin 1978, example (140), page 115)

One might however argue that in (6a), the PP is not a clear modifier but may be an argument. The two examples in (7) from German (from the RRGparbank), where extraposed relative clauses are more frequent, are two cases where the antecedent NP is part of a PP that is clearly a modifier, i.e., a periphery PP. The same holds for (1b) above.

(7) a. [...] über die sie [mit einem unumwunden höhnischen
[...] about which she with an outright mocking
Hass]$_{PP_{peri}}$ sprach, [*der Winston ganz unsicher machte*] [...]
hatred talked that Winston quite uneasy made [...]
'about which she talked with an outright mocking hatred that made Winston quite uneasy'[12]

b. [...] dass der Tod seiner Mutter [...] [auf eine *Weise*]$_{PP_{peri}}$
[...] that the death of.his mother [...] in a *way*
traurig und tragisch gewesen war, [*die es heutzutage*
sad and tragic been had that EXPL these.days
nicht mehr gab].
not more existed
' [...] that the death of his mother had been sad and tragic in a way that did not exist any longer these days.'[13]

Depending on the way PPs are decomposed into elementary building blocks, the link is still relatively close. But the antecedent noun can also be further embedded. Examples were already given in (1c) and (1d). In addition, (8), cited after Walker (2017), gives naturally occurring examples which Strunk and Snider (2013) have found in English corpora, and (9) gives further German examples (a constructed example from Kiss 2005, and three corpus examples, one from TüBa-D/Z, mentioned by Strunk and Snider 2008, and two from the RRGparbank) of extraposed restrictive relative clauses with an embedded antecedent NP. Note that in the case of (9d), there are two extraposed relative clauses, both with a genitive antecedent NP embedded in another NP. Indices indicate which relative clause modifies which antecedent NP. The second, embedded relative clause is definitely restrictive while the first one is rather non-restrictive.

(8) a. A wreath was placed in [the doorway of [the brick *row-house*]$_{NP}$]$_{NP}$ yesterday, [*which is at the end of a block with other vacant dwellings*]. (Walker 2017, example (18c), p.16, originally from Strunk and Snider 2013)

[12] RRGparbank sentence no. 2441, de, 02 April 2021.
[13] RRGparbank sentence no. 517, de, 02 April 2021.

b. For example, we understand that Ariva buses have won [a number of [contracts for [*routes* in [London]$_{NP}$]$_{NP}$]$_{NP}$]$_{NP}$ recently, [*which will not be run by low floor accessible buses*]. (Walker 2017, example (18d), p.16, originally from Strunk and Snider 2013)

(9) a. Man hat [die Frau [des *Boten*]$_{NP}$]$_{NP}$ beschimpft, [*der den*
one has the wife of.the messenger insulted who the
Befehl überbrachte].
command delivered
'The wife of the messenger who delivered the command was insulted.' (Kiss 2005, example (12), page 4)

b. Und dann sollte ich [Augenzeuge [der Zerstörung [einer
and then should I eye.witness of.the destruction of.a
Stadt]$_{NP}$]$_{NP}$]$_{NP}$ werden, [*die mir am Herzen lag*] –
city become that me to.the heart laid –
Sarajevo
Sarajevo
'And then I was about to become an eye witness of the destruction of a city that was dear to my heart – Sarajevo' (Strunk and Snider 2008, slide 15)[14]

c. Wenn Schauprozesse stattfanden, hatte sie [ihren
when public.trials were.happening had she her
Platz [unter [der *Abordnung* der Jugendliga]$_{NP}$]$_{PP}$]$_{NP}$
place among the detachments of.the Youth.League
eingenommen, [*die* [...] *vor dem Gerichtsgebäude*
taken who [...] in.front.of the courthouse
Stellung bezog [...]].
positions took.up [...]
'When public trials were happening she had taken her place among the detachments from the Youth League who took up positions in front of the courthouse.'[15]

[14] Tübinger Baumbank des Deutschen / Schriftsprache (TüBa-D/Z), sentence 16294.

[15] RRGparbank sentence no. 3125, de, 02 April 2021.

d. Er begann [gehäufte Löffel [des *Eintopfgerichtes*$_1$]$_{NP}$]$_{NP}$
he started heaped spoons of.the stew
herunterzuschlingen, [*in dessen schlüpfriger Masse auch*
swallow in whose slimy mass also
[*Würfel* [*eines schwammigen, rosafarbenen Zeugs*$_2$]$_{NP}$]$_{NP}$
cubes of.some spongy pink stuff
auftauchten, [*das vermutlich ein Kunstfleischprodukt*
appeared which presumably a artificial.meat.product
war]$_2$]$_1$.
was
'He started swallowing spoonfuls of the stew, in whose slimy mass appeared cubes of a spongy pinkish stuff which was presumably an artificial meat product.'[16]

Given the examples in (1c), (1d) and (8) and (9), an antecedent noun for an extraposed relative clause that is deeper embedded should in principle be possible and an analysis has to account for that.

So far, all examples we considered contained only a single extraposed relative clause. This is, however, no strict limitation. Examples with more than one extraposed relative clause within the same clause sometimes are acceptable. Example (10), cited after Walker 2017, shows two sentences where we have two extraposed relative clauses with different antecedent NPs. An example in German (our own example) is given in (11).

(10) a. *Someone*$_1$ picked some *books*$_2$ up [*which were lying on the table*]$_2$ [*who really didn't want to*]$_1$. (Baltin 2006, page 241–242)

b. *No one*$_1$ puts *things*$_2$ in the sink [*that would block it*]$_2$ [*who wants to go on being a friend of mine*]$_1$. (Fodor 1978, page 452)

(11) *Keiner*$_1$ wird *die*$_2$ verraten, [*die nicht jubeln*]$_2$, [*der selber*
nobody will those betray who not cheer who himself
am Regime zweifelt]$_1$.
the regime doubts
'Nobody who doubts the regime himself will betray those who don't cheer.'

[16] RRGparbank sentence no. 869, de, 05 April 2021.

However, only in certain cases are such multiple extraposed relative clauses acceptable. It might be that the mirroring property mentioned for instance in de Vries (2002) plays an important role, which states that the antecedent–relative clause pairs must have a nested order. That is, if NP_2 follows NP_1, and both are modified by extraposed relative clauses, then the one modifying NP_2 must precede the one modifying NP_1. But this is probably not the only factor responsible for the unacceptability of certain examples. In the examples in (10) and (11), the second, outermost relative clause has pronominal antecedents such as *no one* or *someone*; it seems to be more restricted with respect to the possible antecedent NPs, and the focus structure might play a role.

In this paper, we aim at allowing in principle for multiple extraposed relative clauses with different antecedent NPs but we do not model restrictions on their order in a general way. Within the analysis we propose, the mirroring property could however be modelled as a restriction on derivation order (see Section 3.4.2 for a brief discussion).

Another possibility is to have several extraposed relative clauses modifying the same noun. Example (12a) gives an example of such stacked relative clauses that are not extraposed, and (12b) (our own example, judged acceptable by several native speakers) gives an example with extraposition.

(12) a. The *theory of light* [*that Newton proposed*] [*that everyone laughed at*] was more accurate than the one that met with instant acceptance. (McCawley 1998, example 3c, page 382)

b. He explained the *theory of light* to her [*that Newton proposed*] [*that everyone laughed at at the time*].

Note, however, that only the first relative clause clearly is a restrictive relative clause. The second is rather non-restrictive. In contrast, in the following examples (13) and (14), we have several restrictive relative clauses.

Example (13a) is an example of stacked relative clauses from RRGparbank. A variant of this with extraposed relative clauses (our own example) is in (13b), again judged acceptable by several native speakers.

(13) a. After confessing to these things they had been pardoned, reinstated in the Party, and given *posts* [*which were in fact sinecures*] [*but which sounded important*].[17]

b. After confessing to these things, *posts* were given to them [*which were in fact sinecures*] [*but which sounded important*].

Concerning German, where (due to the verb-final word order) extraposed relative clauses are more frequent, we found such examples in RRGparbank, see (14).[18]

(14) a. Unzählige Male hatte sie [...] [die Hinrichtung [von
numerous times had she [...] the execution of
[*Menschen*]$_{NP}$]$_{PP}$]$_{NP}$ gefordert, [*deren Namen sie nie*
people demanded whose names she never
zu.vor gehört hatte] [*und an deren angebliche Verbrechen*
before heard had and in whose alleged crimes
sie nicht im entferntesten glaubte].
she not in the.least believed
'On numerous occasions, she had [...] demanded the execution of people whose names she had never heard before and in whose alleged crimes she did not even remotely believe.'[19]

b. [...] wie sie [auf [das Vorbeikommen [von
[...] how they for the passing of
[*Lastautos*]$_{NP}$]$_{PP}$]$_{NP}$]$_{PP}$ gewartet hatten, [*die gewisse*
trucks waited had which certain
Fernfahrten machten] [*und von denen man*
long.distance.journeys made and of which one
wusste, dass sie Viehfutter geladen hatten]; [...]
knew that they cattle.feed loaded had [...]
'[...] how they had waited for trucks to pass, which made certain long distance journeys and which were known to be carrying cattle feed; [...]'[20]

[17] RRGparbank sentence no. 1376, en, 12 Feb 2021.

[18] The fact that "zu vor" appears in the first sentence as two tokens (instead of one, which would have been correct) is a tokenization error in the electronic version. In the original text, it is one word (Orwell 2000, page 141, l. 6–7).

[19] RRGparbank sentence no. 3124, de, 12 Feb 2021.

[20] RRGparbank sentence no. 3301, de, 12 Feb 2021.

Note that the two examples in (14) are not only examples of multiple extraposed relative clauses but, in addition, display cases of embedded antecedent NPs since in both cases, the antecedent is embedded within an argument NP (resp. PP) of the matrix verb.

One might argue that in these examples of multiple extraposed relative clauses, only a single clause has been extraposed consisting of a coordination of two relative clauses. In RRG, however, two clauses that are coordinated and that (can) have different tense values, form a SENTENCE; since in a CLAUSE cosubordination, i.e., a CLAUSE with two CLAUSE daughters, the two clauses share certain features, such as tense. But, on the other hand, restrictive relative clauses are assumed to be of category CLAUSE. Therefore, the standard RRG analysis would tend to assume multiple relative clauses in these cases, as well as in the extraposed case as in (14). This is also in line with the annotations we find in the RRGparbank.

Besides this rather theory-internal argument, a further point in favour of assuming two different relative clauses instead of a complex one is that we can also have cases where only one of the two relative clauses is extraposed, as in (15). Of course, in this case, neither needs a conjunction but that can be modelled via appropriate features.

(15) [...] wie sie [auf [das Vorbeikommen [von [*Lastautos* [*die*
[...] how they for the passing of trucks which
gewisse Fernfahrten machten]]$_{NP}$]$_{PP}$]$_{NP}$]$_{PP}$ gewartet
certain long.distance.journeys made waited
hatten, [*von denen man wusste, dass sie Viehfutter geladen*
had of which one knew that they cattle.feed loaded
hatten] ; [...]
had [...]
'[...] how they had waited for trucks to pass, which made certain long distance journeys and which were known to be carrying cattle feed; [...]'

We therefore assume that the sentences in (14) are cases of multiple extraposed relative clauses that do not form a single complex extraposed relative clause. As already mentioned, the fact that the second one needs a conjunction can be captured via some appropriate feature that enforces the adjunction of the clause linkage marker.

An additional complication arises from the fact that some determiners, such as *derjenige* ('the one') in German, require a relative clause (Alexiadou *et al.* 2000; Sternefeld 2008). Examples (16a) and (16b) are grammatical while (16c) is not. In German, the relative clause in this case can be adjacent to its antecedent or extraposed. *Derjenige* used as a pronoun, i.e., without a noun, behaves exactly the same way.

(16) a. *Derjenige (Läufer)*, [*der zuerst ins Ziel läuft*], gewinnt.
the.one (runner) who at.first into.the goal runs wins
'The runner who finishes first wins.'

b. *Derjenige (Läufer)* gewinnt, [*der zuerst ins Ziel läuft*].

c. *Derjenige (Läufer) gewinnt.

The following examples (17) are actual corpus examples with an antecedent NP *diejenigen* ('those') and an extraposed relative clause, taken from the German part of RRGparbank. In both cases, the relative clause is obligatory.

(17) a. In gewisser Weise ließen sich *diejenigen* am
in certain way let themselves those most
leichtesten von der Parteidoktrin überzeugen, [*die ganz*
easily of the Party.doctrine convince who totally
außerstande waren, sie zu verstehen].
incapable were it to understand
'In a way, those who were totally incapable of understanding it, could most easily be convinced of the Party doctrine.'[21]

b. [...] *diejenigen* zu notieren und verschwinden zu lassen,
[...] those to mark.down and disappear to let
[*die vielleicht gefährlich werden konnten*]
who perhaps dangerous become might
'to mark down and eliminate those who might potentially become dangerous'[22]

Note that the requirement for a restrictive relative clause is actually rather a requirement for some additional specification that could also

[21] RRGparbank sentence no. 3203, de, 12 Feb 2021.
[22] RRGparbank sentence no. 1317, de, 12 Feb 2021.

be met by something other than a relative clause, for instance a genitive NP:

(18) In einer von D. verfaßten Denkschrift sind alle seine
in a by D. written memorandum are all his
Bauten und diejenigen seiner zahlreichen Schüler
buildings and the.ones of.his numerous pupils
verzeichnet.
listed
'In a memorandum written by D. all his buildings and those of his numerous pupils are listed.'[23]

In such a case, the request for additional information would already be satisfied at the NP node, due to the adjunction of the NP *seiner zahlreichen Schüler* ('of his numerous pupils').

So far, we have concerned ourselves with data showing how non-local the phenomenon of extraposed relative clauses is. There are, however, also limitations on how far apart from each other the relative clause and its antecedent can be. One is the *Right Roof Constraint* (Ross 1967), stating that no maximal projection can be in between the antecedent NP and the clause that the relative clause attaches to (see for instance Crysmann 2013). Examples in (19) (our own examples) illustrate this; further examples can be found in Ross (1967).

(19) a. [*diejenigen* zu notieren]$_{CORE}$ hat er versprochen, [*die*
those to mark.down has he promised, who
vielleicht gefährlich werden konnten]
perhaps dangerous become might
'He has promised to mark down those who might potentially become dangerous'

b. er hat versprochen, [dass er *diejenigen* notiert, [*die*
he has promised that he those marks.down who
vielleicht gefährlich werden konnten]]$_{CLAUSE}$
perhaps dangerous become might

[23] In: Olbrich, Harald (ed.), Lexikon der Kunst, Berlin: Directmedia Publ. 2001 [1989], page 6533. From the Kernkorpus of the Digitales Wörterbuchs der deutschen Sprache, `https://www.dwds.de/d/korpora/kern`, 09 April 2021.

'He has promised that he marks down those who might potentially become dangerous'

c. *[dass er *diejenigen* notiert]$_{CLAUSE}$ hat er versprochen, [*die vielleicht gefährlich werden konnten*]

Concerning extraposed relative clauses in German, it has been proposed that the antecedent NPs cannot be in the vorfeld of the sentence (see the discussion in Holler 2013, page 271), i.e., preceding the finite verb when the finite verb is in second and not in final position. Examples are in (20).

(20) a. *Der Mann hat die Frau getroffen, der im Kino
the man has the woman met who at.the cinema
war.
was
'The man who was at the cinema has met the woman.'[24]

b. ?Dem Mann hat sie etwas zugeflüstert, der dort
the man has she something to-whispered who there
steht.
stands
'She whispered something to the man who is standing there.'
(example (47), p.24, Büring and Hartmann 1997)

Note, however, that (20b), though so marked, is, according to Büring and Hartmann (1997), not ungrammatical. As observed also by Holler (2013), it seems that a contrastive focus on the vorfeld constituent makes such examples much better. Concerning (20a), imagine for instance a situation where there are three men, one went to the cinema, one to the theater and the third one to a concert. And we know that one of them met the woman we are interested in. In that case the following dialogue is perfectly fine:

(21) Welcher der drei Männer hat nochmal die Frau
Which of.the three men has again the woman
getroffen? – DER *Mann* hat die Frau getroffen, [*der im*
met? – the man has the woman met who at.the
Kino war].
cinema was

[24] Example provided by an anonymous reviewer.

‘Which of the three men met the woman again? – The man who was at the cinema met the woman.’

Other examples where the NP in the vorfeld is a perfect antecedent for the extraposed restrictive relative clause are the ones in (22).

(22) a. *Jeder* wird dieses Lied sofort wiedererkennen,
everybody will this song immediately recognize
[*der es schon einmal gesungen hat*].
who it already once sung has
‘Everybody who has already sung this song once will recognize it immediately.’ (our own example)

b. Nur die *Wanderer* waren erschöpft, [*die den Gipfel*
only those hikers were exhausted who the summit
erklommen hatten].
climbed-to had
‘Only those hikers were exhausted who had climbed to the summit.’ (Holler 2013, example (30), page 276)

c. Der fette *Musiker* von Achselroths Tisch kam herein
the fat musician from Achselroth’s table came in
[*der schon einmal bis Kuba gekommen war*].
who already once as.far.as Cuba come had
‘The fat musician from Achselroth’s table came in who had already come as far as Cuba once.’[25]

This paper is not concerned with modelling focus, which would be necessary in order to capture the (in)acceptability of sentences with a vorfeld antecedent for an extraposed restrictive relative clause. Given the preceding examples, we choose to allow any NP, whether in the vorfeld or mittelfeld, to serve in principle as antecedent to extraposed relative clauses.

Concerning the structure of the relative phrase, it can also be the case that the relative pronoun is not an argument of the verbal head of the relative clause. Two examples from RRGbank (i.e.,

[25] In Seghers, Anna: Transit, Gütersloh: Bertelsmann 1995 [1943], page 175. From the Kernkorpus of the Digitales Wörterbuchs der deutschen Sprache, `https://www.dwds.de/d/korpora/kern`, 09 April 2021.

from the PTB) are given in (23). In (23b), the relative pronoun is embedded in an argument and, futhermore, parts of the argument NP ("49% of which") is stranded, i.e., is positioned inside the CORE.

(23) a. He assumed the missing piece contained a *gene or genes* [*whose loss had a critical role in setting off the cancer*].[26]

b. It said the programs , largely game shows , will be provided by its E.C. Television unit along with Fremantle International , a *producer and distributor of game shows* [*of which it recently bought 49 %*].[27]

Such cases, where the relative pronoun is embedded in an argument of the head of the relative clause, can occur in combination with extraposition. A German example from Wikipedia is given in (24), and an example from the German part of RRGparbank is given in (25).

(24) Räuberschach ist eine Schachvariante, bei der
robber.chess is a chess.variant in which
Schlagzwang besteht und derjenige *Spieler* gewinnt,
capturing.obligation holds and the.one player wins
dessen Spielsteine alle geschlagen wurden.
whose pieces all captured have.been
'Robber Chess is a chess variant in which capturing is obligatory and the player whose pieces have all been captured is the winner.' (Wikipedia)[28]

(25) Die ungewöhnliche Anlage des Zimmers war zum Teil für
the unusual setting of.the room was partly for
den *Gedanken* verantwortlich, [*zu dessen Verwirklichung er*
the thought responsible to whose realization he
jetzt schritt].
now went

[26] RRGbank sentence no. 9028, 13 Feb 2021.
[27] RRGbank sentence no. 1153, 13 Feb 2021.
[28] https://de.wikipedia.org/wiki/Räuberschach, 05 Nov 2019.

'The unusual geography of the room was partly responsible for the idea that he was now about to realize.'[29]

Besides these cases of complex relative phrases, we can also have a long-distance dependency within the relative clauses such that the relative pronoun is an argument of an embedded verb. An example from the English part of RRGparbank is given in (26) where the relative pronoun is an argument of the embedded predication *to take a look at*.

(26) "There 's another *room* upstairs [*that you might care to take a look at*]," he said.[30]

In this paper, we will concentrate on establishing the relation between relative clause and antecedent NP, and we will leave the cases exemplified in (23)–(26) aside, given that the phenomena in these sentences are to a large extent independent from the difficulty of linking extraposed relative clauses to their antecedents.

2.3 *Formalizing Role and Reference Grammar: Tree Wrapping Grammar*

In the following, we adopt the formalization of RRG as a tree rewriting grammar, more precisely a TWG (Kallmeyer *et al.* 2013; Kallmeyer 2016; Kallmeyer and Osswald 2017; Osswald and Kallmeyer 2018). A TWG consists of a finite set of *elementary trees* that can be combined into larger trees via *substitution*, *sister adjunction* and *wrapping substitution*. Substitution simply replaces a non-terminal leaf (called a *substitution node*) with a new tree, provided the category of the substitution node and the root category of the new tree are the same and the new tree is not an adjunct tree. Sister adjunction adds a new adjunct tree to a node, provided that the category of the root of the newly added tree and the category of the adjunction site are the same. Adjunct trees are such that the root is marked with an asterisk and below the root, there is only a single daughter tree. This new daughter tree can be inserted at any

[29] RRGparbank sentence no. 81, de, 12 Feb 2021.
[30] RRGparbank sentence no. 1853, en, 12 Feb 2021.

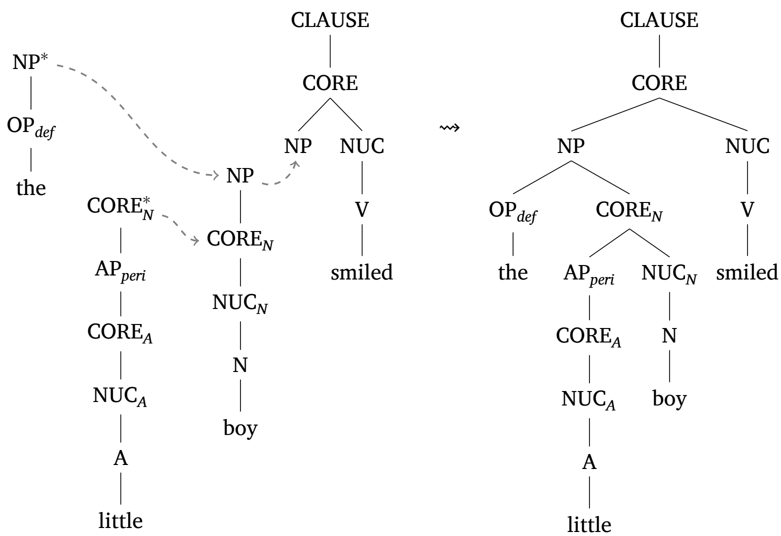


Figure 5: Example involving substitution (argument filling) and operator/periphery adjunction

position among the other daughter subtrees below the adjunction site.

Roughly, substitution is used to add arguments while sister adjunction is used to add operators and periphery elements. A sample derivation involving one argument insertion (substitution), one operator adjunction and one modifier (i.e., periphery element) adjunction is given in Figure 5.

The third operation, wrapping substitution, is the one that adds expressive power to the formalism. It adds a tree with a *d-edge* (= dominance edge) between a node v_1 and its *d-daughter* v_d to a derived tree that has a substitution node with the same category as v_d and an internal node v (which can be the root) with the same category as v_1. The substitution node is replaced with the subtree below v_d while the node v_1 merges with the node v of the target tree, thereby adding new daughter trees to v (to the left or to the right of the already existing daughters) or new nodes dominating v (the latter is only allowed if v is the root).[31] Wrapping substitution is used for extraction; the filling of the substitution node adds an argument while the upper part adds

[31] Note that this is the slightly relaxed definition of wrapping from Bladier *et al.* (2020a).

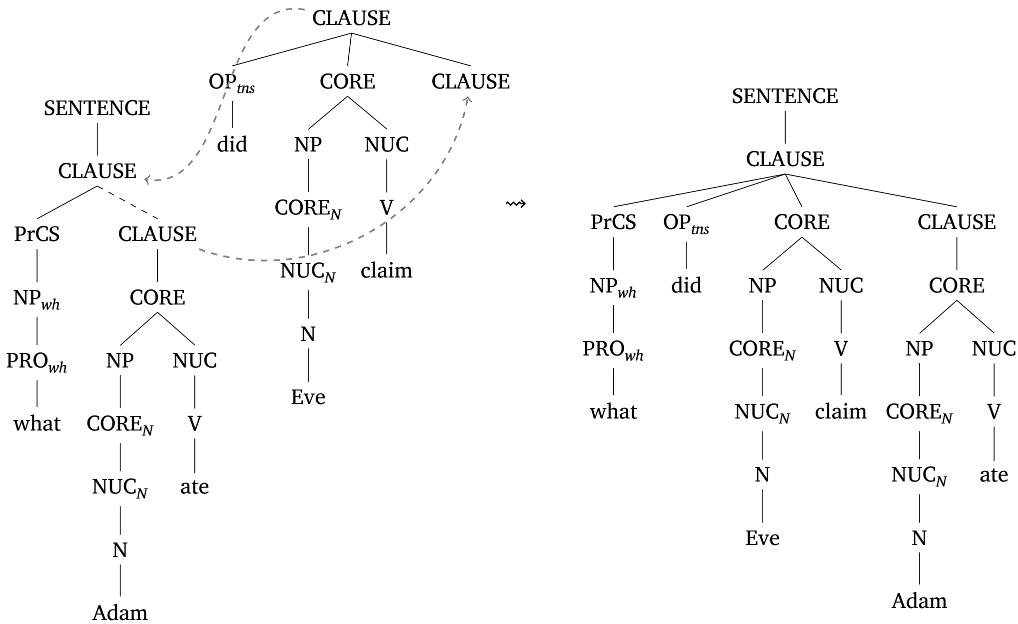


Figure 6: Sample wrapping substitution

material that is extracted out of that argument. A sample wrapping substitution is shown in Figure 6.

As in the case of TAG, nodes can have features, though not bottom and top feature structures but just a single feature structure. As in TAG, feature structures are untyped and restricted in depth such that only a finite set of feature structures is possible. Besides nodes, edges can have left and right features, expressing what is expected to the left/right of a node respectively. We will introduce these features and the way they unify more in detail in Section 4.1.

Note that TWG does not allow for crossing branches, i.e., cannot yield exactly the trees we find in the RRGbank. See for instance Figure 1, where the tense operator *will* attaches at the CLAUSE node, which leads to a crossing branch. Put differently, the yield of the CORE node has a gap. The TWG formalization would attach the tense operator lower while capturing the fact that it scopes at CLAUSE level in the features (see Kallmeyer and Osswald 2017 for more details).

TWG is more powerful than TAG (Kallmeyer 2016). There are two main reasons: a) TWG allows for more than one wrapping sub-

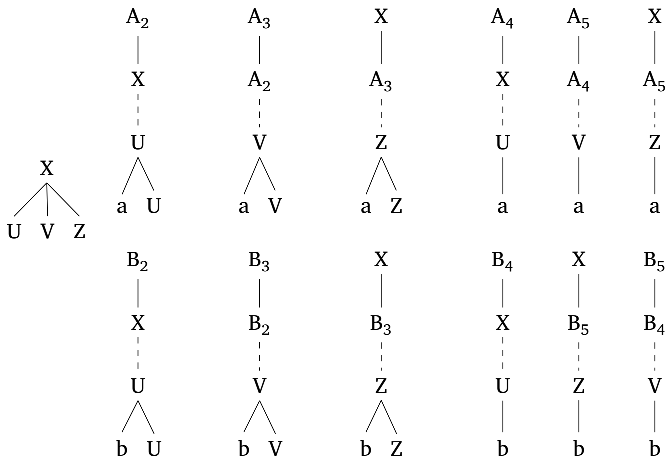


Figure 7: TWG for the double copy language $\{w^3 \mid w \in \{a,b\}^+\}$

stitution stretching across specific nodes in the derived tree; and b) the two target nodes of a wrapping substitution (the substitution node and the higher internal node) need not come from the same elementary tree, which makes wrapping non-local compared to adjunction in TAG. To see why this property matters, consider the sample TWG in Figure 7, which generates the double copy language, a language that is not a tree adjoining language. The idea of this TWG simply is that the new *a*'s (respectively *b*'s) for the three copies are added one after the other from left to right, and the root label always determines which substitution slot has to be filled next. Root label X means that U has to be filled next, root label A_2 (respectively B_2) means that V has to be filled next, and so on. Figure 8 shows a sample derivation with this grammar.

If the number of d-edges that stretch across a certain node and that are not nested within each other is limited to some *k* (this type of TWG is called *k*-TWG), one can show that for every *k*-TWG, a simple Context-Free Tree Grammar (CFTG, Kanazawa 2016) of rank *k* can be constructed (Kallmeyer 2016). Simple CFTGs of rank *k* are, in turn, equivalent to well-nested Linear Context-Free Rewriting Systems (LCFRS) of fan-out $k+1$. Consequently, 1-TWGs are weakly equivalent to TAG while *k*-TWGs in general are more powerful. The TWG in Figure 7 is a 3-TWG.

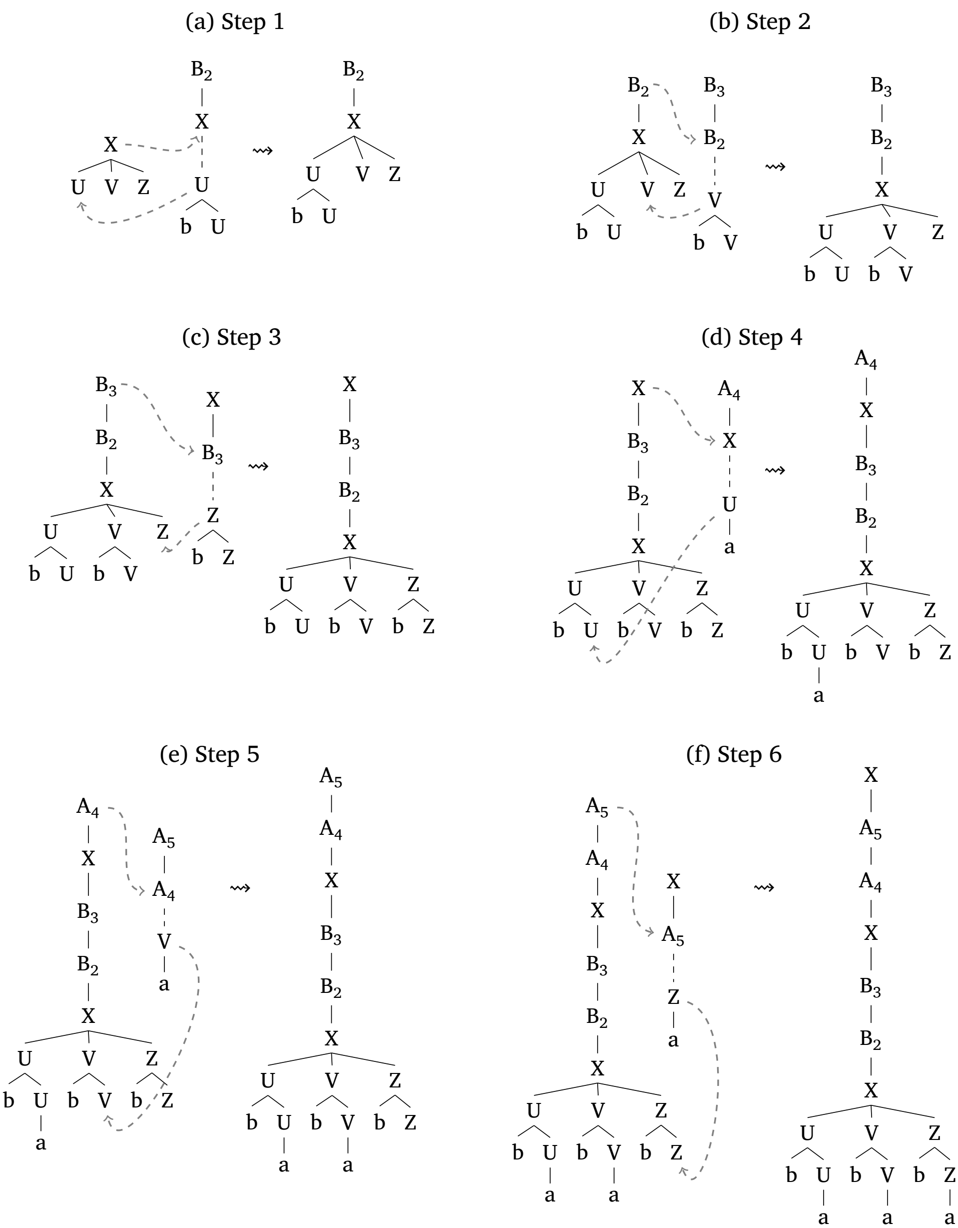


Figure 8: Sample TWG derivation for *bababa*, for the TWG from Figure 7

ANALYSIS 3

Given the trees RRG assumes for relative clauses and given the TWG formalization, we will now address the question of how the underlying elementary trees could look and how they might combine. Before coming to extraposed relative clauses, let us start by giving an analysis of (4a) (repeated here as (27a)), where the relative clause is adjacent to its antecedent noun. In this case, we can simply add it as a further daughter to the NUC_N node using sister adjunction. This is our standard way of adding peripheral elements, i.e., modifiers. Figure 9 gives the corresponding adjunction step.

(27) a. A *girl* [*who was singing a song*] came in.
b. A *girl* came in [*who was singing a song*].

We assume the following features in order to capture the type and scope of the relative clause: The CLAUSE node of the relative clause has a feature PERI that characterizes the type of category that this relative clause modifies (i.e., of which it is a periphery element). Here, we have a restrictive relative clause, which means that it modifies the nucleus of an NP (which is of category NUC_N). Furthermore, a second feature PERI-SCOPE has as its value the identifier of the relevant NUC_N node. Node identifiers are captured within a feature N(ODE)-ID. This feature is used to pass the id of the antecedent NUC_N node into the PERI-SCOPE value of the relative clause via unification at the sister adjunction site (here unification of $\boxed{1}$ and $\boxed{2}$).

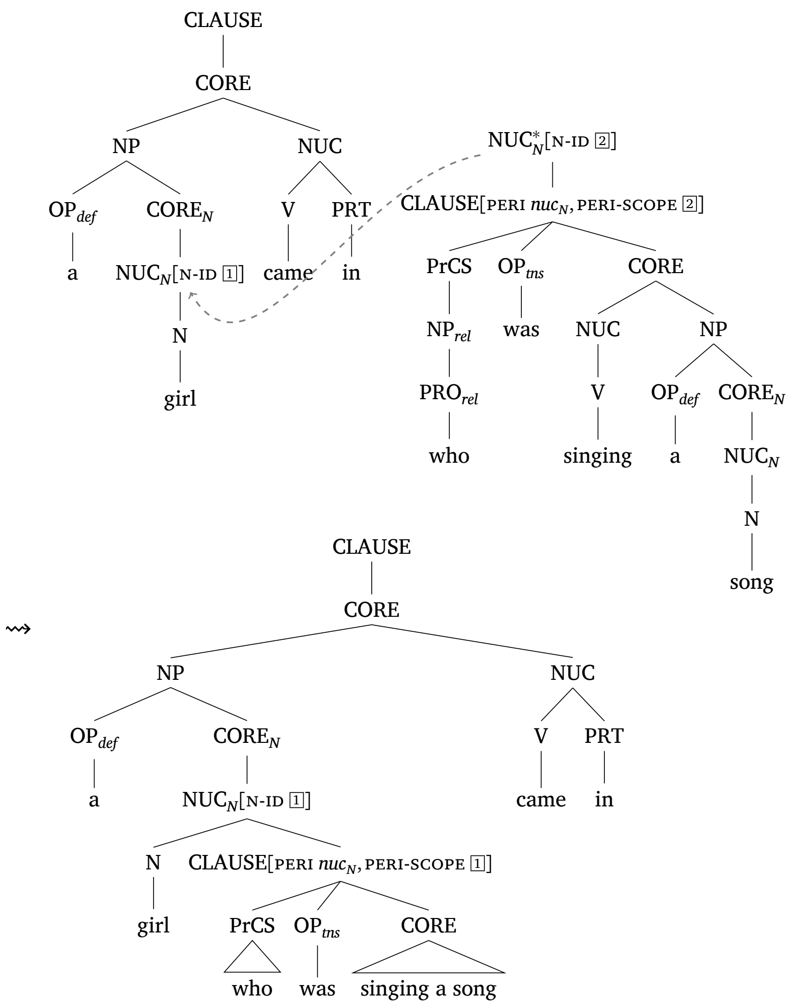
This analysis yields the structures standardly assumed for restrictive relative clauses in RRG (see Van Valin 2005, and see Section 2.1 above) and it is in line with the proposal to use sister adjunction for adding periphery elements (see Kallmeyer *et al.* 2013).

To our knowledge, extraposed relative clauses have not been analyzed in RRG and there is no proposal for extraposed relative clauses in any tree rewriting grammar (such as TAG or variants of TAG). In the following, we will develop and discuss three options.

Approach 1: Anaphoric approach 3.1

Now let us go through several possibilities for analysing (4b) (repeated above as (27b)), i.e., the variation of (4a) with the relative clause be-

Figure 9: RRG analysis of (4a) with relative clause as periphery



ing extraposed. One possibility is to add the relative clause by sister adjunction to the verbal CLAUSE node and to mark the fact that this is a nominal NUC_N periphery element within the features, as in Figure 10. Within the derived constituency tree, the features of the relative clause tell us that this is a periphery element of some nominal nucleus but they do not specify which NUC_N is the antecedent. In this respect, it contains less information than the structure derived for a non-extraposed relative clause as in Figure 9. In order to find the antecedent, one would have to find an appropriate NUC_N node in one of

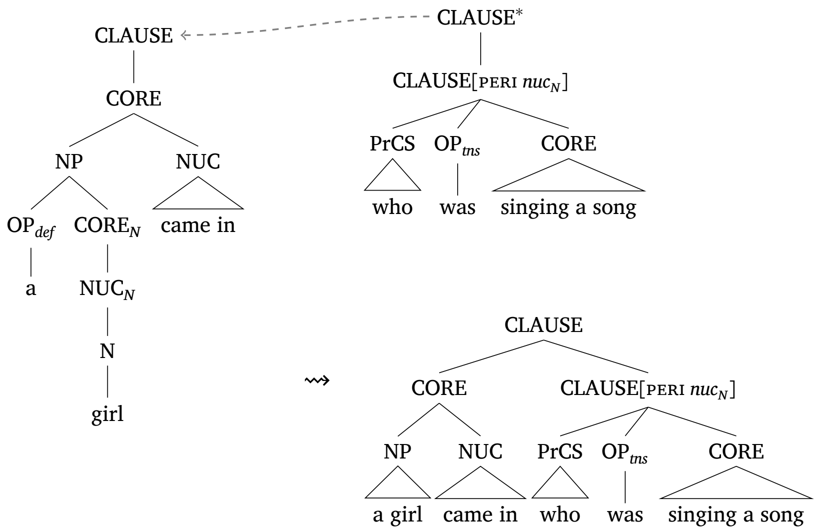


Figure 10: First possibility for (4b) (no link to antecedent NUC_N)

the sisters of the relative clause tree. This resolution step would be a separate post-processing step.

This first approach is in a sense an anaphoric approach since the linking of relative clause to antecedent is considered an anaphoric link that is established by some non-local process that operates on the derived tree. It is close to what Kiss (2005) proposes as a 'semantic' approach where, syntactically, relative clause extraposition is considered as ordinary adjunction, and the link to the antecedent is established via a condition on interpretation (his condition (16), page 7), which states that a suitable antecedent has to be found in the phrase to which the relative clause adjoins. This search for a suitable antecedent can be realized in HPSG via appropriate principles (see the related work in Section 5.2). An anaphoric approach within a tree rewriting grammar such as TAG or TWG would, however, have difficulties capturing syntactic constraints that are due to the syntactic dependency between antecedent NP and relative clause. For instance, agreement between antecedent noun and relative pronoun is not accounted for or, rather, has to be accounted for when resolving the anaphoric link to an antecedent noun, along the lines of pronoun antecedent resolution, which is a choice one can make. More problematic is, however, that there are NPs for which the adjunction of a (possibly extraposed) relative clause is obligatory. Examples are the above-mentioned classes of determiners, such as *derjenige* in German (see Section 2.2). This is something that one might want to capture within syntax and not in a separate module of anaphora resolution. But an approach that does not establish a syntactic link between antecedent noun and relative clause cannot do so.

3.2 *Approach 2: NPs provide landing sites for relative clauses*

The following two options both assume that the antecedent NP and the CLAUSE to which the relative clause attaches are part of the same elementary tree, linked by a d-edge. One such possibility is to have this d-edge in the tree of the antecedent, i.e., add the NP that is modified by the relative clause via wrapping substitution. The upper part of its elementary structure could be a slightly degenerate single node that adds only an identifier. That is, the upper part is a CLAUSE node that identifies with the root of the verbal tree. It adds an identifier of the embedded nominal NUC node in order to provide access to it when adding a modifier. To this end, we use a feature NUC-N-ID (for NUC_N node identifier) on the upper CLAUSE node. The step of adding the NP *a girl* is shown in Figure 11a while Figure 11b gives the subsequent step where the extraposed relative clause is added. It is adjoined to the CLAUSE node but retrieves its antecedent (feature PERI-SCOPE) via the NUC-N-ID on the CLAUSE node, which is the N-ID feature from the antecedent NUC_N node. These two steps of wrapping substitution and adjunction could also be performed in reverse order, i.e., first adjoining the relative clause to the root of the NP tree and then wrapping the NP tree around its predicate. The result would be the same, and, furthermore, the derivation would of course also be the same since the way the elementary trees combine are identical.

Note that this analysis allows also for more embedded antecedent NPs (as in the examples in (8)), as long as they are added by filling a substitution slot. This is due to the non-locality of the wrapping operation: When wrapping a tree γ_1 around some tree γ_2, the upper part of γ_1 targets some internal node of γ_2, no matter whether this internal node and the substitution node in γ_2 that gets filled come from the same elementary tree or not.

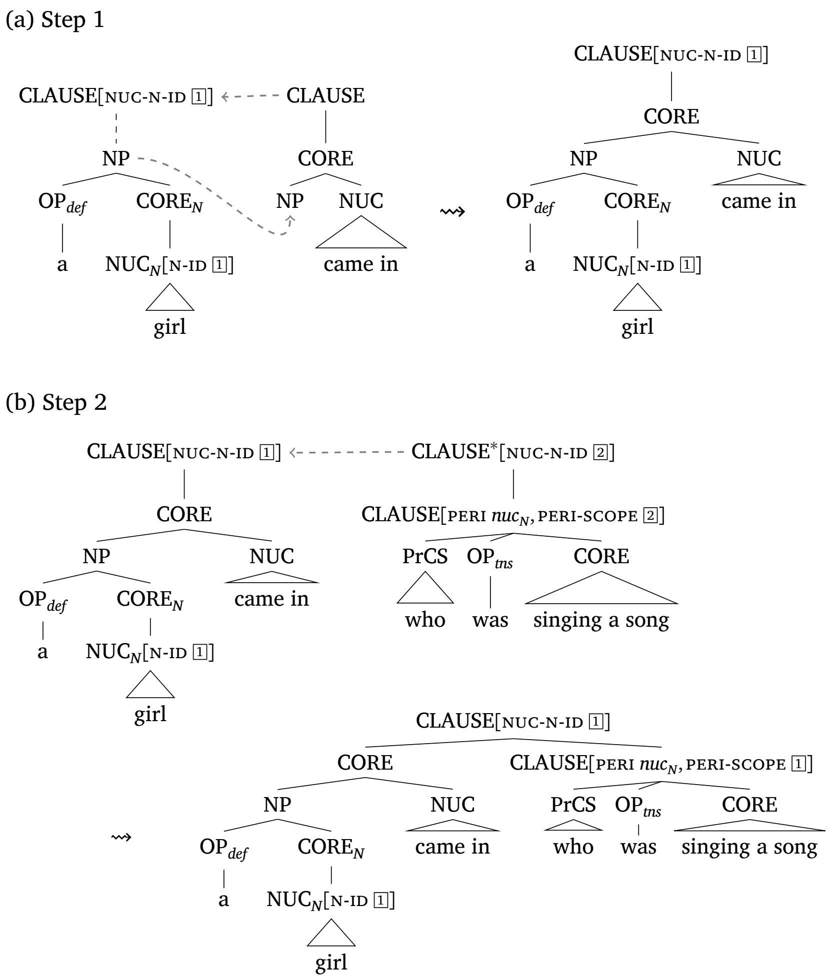


Figure 11: Second possibility for (4b): the antecedent provides a "landing site" for extraposed relative clauses

When adopting such an analysis, we need to make sure that at most one NP below a CLAUSE node provides such a node identifier for attaching an extraposed relative clause. Take for instance (28). With trees along the lines of Figure 11a for both NPs (*a girl* with N-ID = $\boxed{1}$ and *the room* with N-ID = $\boxed{2}$), we would end up unifying NUC-N-ID = $\boxed{1}$ and NUC-N-ID = $\boxed{2}$ at the CLAUSE node, which would be perfectly possible. The extraposed relative clause would then identify both NUC_N nodes as its scope.

(28) A *girl* entered the room [*who was singing a song*].

This can be avoided. Instead of using variables (which can unify with each other) as in Figure 11a, we use actual labels drawn from a set of node identifiers as values of the features N-ID and NUC-N-ID in the NP trees, for instance *node_1*, *node_2*, ... In our feature structure signature, these values would be part of the set of possible attribute values while the variables we use here are of course not part of the feature structure signature. Each nominal nucleus has then its own unique identifier as value of its N-ID attribute that cannot unify with the (different) identifier of the nucleus of a different NP, being different values of the same attribute. In the case of Figure 11a, we might replace the variable [1] with the attribute value *node_1*. In a sentence with more than one NP, such as (28), we might assume that the first NP (*a girl*) has N-ID = *node_1*, the second (*the room*) has N-ID = *node_2*. If we use trees that provide landing sites for extraposed relative clauses for both NPs, we would have a unification failure at the NUC-N-ID attribute at the respective CLAUSE nodes. Therefore, at most one of them could provide such a landing site.

The second approach comes with the inconvenience that, for each NP, we need an extra elementary structure that is used only for modification of the nucleus with some extraposed relative clause. This is possible but slightly unsatisfying given that the relative clause is a true modifier and should therefore not be anticipated in the elementary structure of the noun. Furthermore, it would lead to spurious ambiguities since such a specialized elementary structure can also be used in cases where no extraposed relative clause is adjoined. An advantage of the second approach might be that it is able to express the fact that for certain NPs such as German *derjenige (N)*,[32] the NP comes with the desire to be modified by a restrictive relative clause. In this case, one would provide only the elementary structure with the single CLAUSE node for attaching a relative clause. Note however, that this does not yet require the adjunction of a relative clause. We will discuss ways to impose obligatory adjunction in these cases below.

A more serious problem with this second approach is that multiple extraposed relative clauses with different antecedent NPs (see (10) above) are not possible since only one NP can provide its NUC-ID as

[32] *Derjenige (N)* stands for either a pronoun *derjenige* or an NP of the form *derjenige N* as for instance *derjenige Läufer* in example (16), page 241.

a feature at the clause node. The restricted form of feature structures used in TWG does not allow list-valued features, as in HPSG.

Approach 3: Relative clauses incorporate their antecedent NPs 3.3

A third and, as we will see, better possibility is to include the d-edge between the CLAUSE node and the NP node in the elementary tree of the relative clause. The NP node can be a leaf node, i.e., a substitution site, that can be filled by the antecedent NP. The combination of antecedent NP and relative clause is a substitution step while the resulting structure is added to the matrix sentence by wrapping substitution. The first step for our example, i.e., combining the NP *a girl* and the relative clause into a complex NP, is given in Figure 12a, and the subsequent step of filling the argument slot of *came in* via wrapping substitution is depicted in Figure 12b. Note that this order is not obligatory; one can also first wrap the relative clause tree around the matrix clause tree (in this case the lower NP leaf merges with the NP substitution slot) and then add the antecedent NP by substitution.

This solution, in contrast to the preceding one, has the advantage that we do not need a special NP tree with a single CLAUSE node, just for the possibility to be modified by an extraposed relative clause. Instead, the NP trees look the same, whether we add a relative clause or not. Furthermore, the problem of accidentally unifying the NUC-ID features of different NPs does not arise since these features do not appear on the CLAUSE node of the matrix sentence, only on the CLAUSE node of the relative clause (feature PERI-SCOPE).

The fact that in this third approach, the combination of extraposed relative clause and antecedent NP is achieved via substitution reflects nicely the semantic argument status of this antecedent NP with respect to the relative clause: In the underlying semantic logical structure, it is either an argument of the predicate denoted by the head of the relative clause or an argument of something embedded within the relative clause.

The third option, in contrast to the second, allows easily for more than one extraposed relative clause, both with different antecedent NPs or with the same. This is because structure sharing concerning the

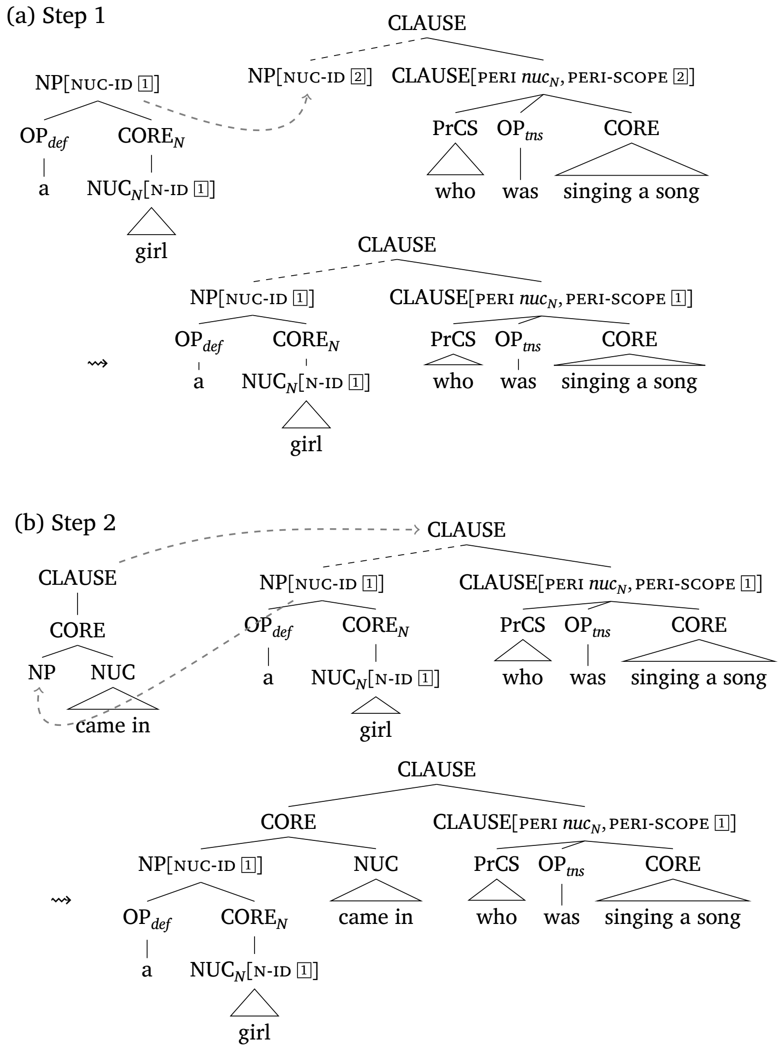


Figure 12: Third possibility for (4b): computing first the complex NP including the extracted relative clause

features NUC-ID and PERI-SCOPE only occurs between the antecedent NP node and the corresponding CLAUSE$_{peri}$ node, and it does not involve the upper CLAUSE node, which might serve as attachment site for multiple relative clauses.

In the following, we adopt the third analysis because it easily covers cases of multiple extraposed relative clauses and it does not require special NP trees that anticipate modification by an extraposed relative clause.

Further constraints on extraposition of relative clauses 3.4

Island constraints 3.4.1

As mentioned above, in between the CLAUSE that the relative clause attaches to and its antecedent NP, no further CLAUSE nodes may appear (see Ross 1967). This could be modelled by excluding certain non-terminal categories on the path spanned by a d-edge, in the spirit of V-TAGs *integrity constraints* (Rambow 1994). In other words, for every d-edge in an elementary tree, we allow the specification of "islands", i.e., of categories that are excluded on the corresponding path. For extraposition of relative clauses, the category CLAUSE would be disallowed. Something similar was proposed in Kallmeyer *et al.* (2013) as a general way to model island constraints in TWG.

Mirroring property 3.4.2

In cases of multiple extraposed relative clauses attaching to the same CLAUSE node, the order of the relative clauses depends on the order in which they are added because each wrapping substitution that fills an NP slot and adds at the same time a corresponding extraposed relative clause, adds this as a new rightmost daughter of the CLAUSE node. If we wanted to restrict the order, for instance according to the mirroring property (de Vries 2002), we could impose a specific derivation order, for instance a filling of argument slots from the right to the left or from the NUC node outwards.

For example in the case of (11), repeated here as (29), we could impose that first the pronoun *die* ('those') is added, which would add the corresponding relative clause as new rightmost element below the CLAUSE. Then, in a subsequent step, one moves to the left and adds *keiner* ('nobody'), which adds the next extraposed relative clause further to the right.

(29) *Keiner*$_1$ wird *die*$_2$ verraten, [*die nicht jubeln*]$_2$, [*der selber*
nobody will those betray who not cheer who himself
am Regime zweifelt]$_1$.
the regime doubts
'Nobody who doubts the regime himself will betray those who don't cheer.'

3.4.3 Agreement between antecedent and relative pronoun

So far, the antecedent NP node and the relative clause share the value of the respective features NUC-ID on the NP and PERI-SCOPE on the CLAUSE node, in order to establish something like a coreference link between the two or, more precisely, to characterize the scope of the relative clause. In addition, we can of course also share other features between the different nodes of the relative clause tree, in particular agreement features. Take for example (30), where the relative pronoun must have agreement features GEN = *n*, NUM = *sg*, which is the reason why *das* is possible while *die* (features either GEN = *f*, NUM = *sg* or NUM = *pl*) yields an ungrammatical sentence (see the second option in (30)).

(30) Das *Team* gewinnt, [*das/*die* *zuerst ankommt*].
the.N team wins that.N/*that.F first arrives
'The team that arrives first wins.'

The relevant derivation is shown in Figure 13. The agreement features (AGR) of the antecedent NP unify with the agreement features of the relative pronoun. This can be achieved via a feature REL-AGR, which is identical with the AGR feature of the relative pronoun. In a case like (30), the latter is also the AGR feature of the whole relative phrase. This is however not always the case. If the relative pronoun is embedded into the NP under PrCS (e.g., *the picture of whom*, *whose daughters*, etc.), the antecedent NP must share its agreement features with the embedded pronoun (transported to the root of the relative phrase via the REL-AGR feature) and not with the entire NP (AGR feature of the relative phrase, variable [2] in our example). The latter plays a role inside the relative clause. In a case of subject relativization like (30) for instance there will be a shared AGR feature between relative NP and the verb of the relative clause (variable [5] in our example). This way of dealing with agreement is very much in line with what we find in TAG, for instance in the XTAG grammar (XTAG Research Group 2001).

4 OBLIGATORY RELATIVE CLAUSES

Now let us turn to the phenomenon that some determiners, such as *derjenige* in German, require a relative clause, see (16), repeated here

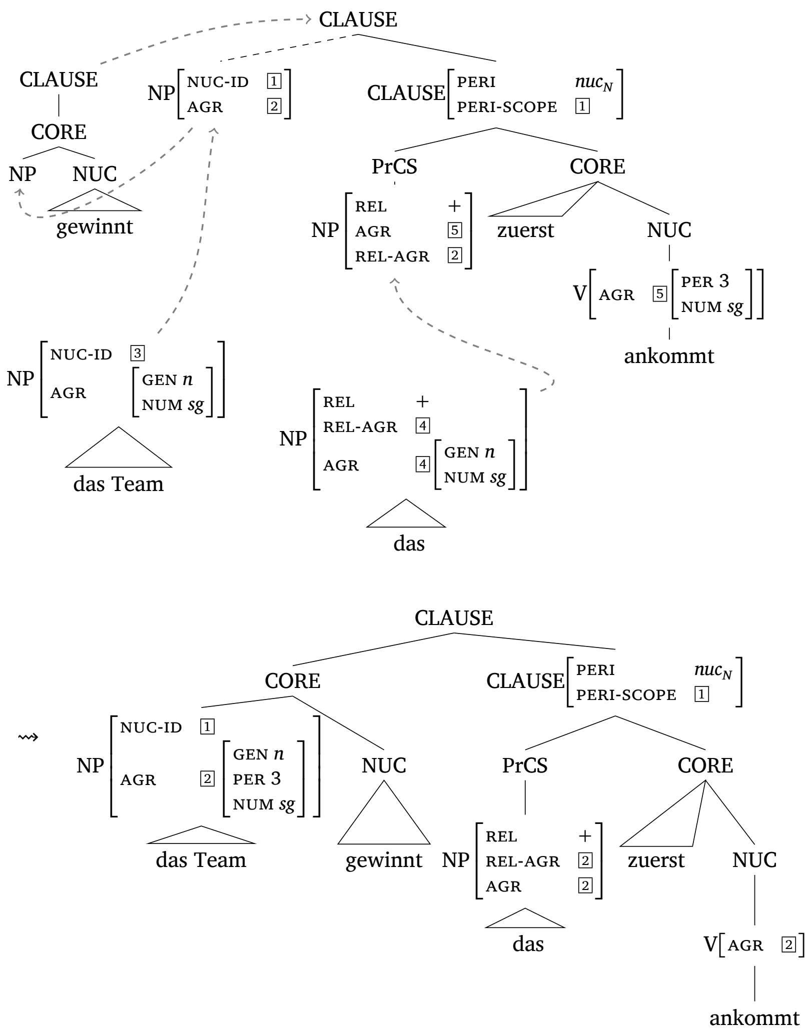


Figure 13: Sharing agreement features between antecedent NP and relative pronoun

as (31). Sentences (31a) and (31b) are grammatical while (31c) is not. The relative clause can be adjacent to its antecedent or extraposed. *Derjenige* used as a pronoun, i.e., without a noun, behaves exactly the same way.

(31) a. Derjenige (Läufer), der zuerst ins Ziel läuft, gewinnt.
the.one (runner) who at.first into.the goal runs wins.
'The runner who finishes first wins.'

b. Derjenige (Läufer) gewinnt, der zuerst ins Ziel läuft.

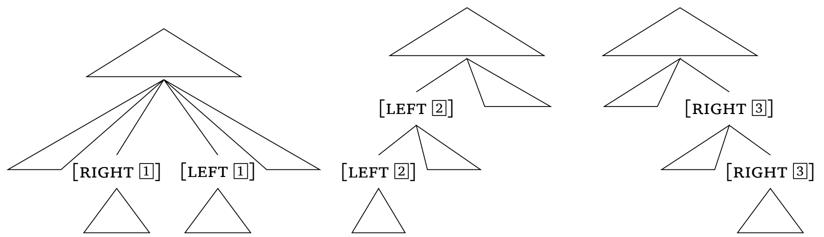
c. *Derjenige (Läufer) gewinnt.

The difficulty is that we want to express an obligatory adjunction constraint. Within the RRG formalization used in this paper, this is usually done via edge features (Kallmeyer and Osswald 2017) that are shared between neighbouring edges and between edges dominating each other via some automatic feature unification mechanism on the final derived tree. But edge features, as defined in Kallmeyer and Osswald (2017), cannot be shared across substitution nodes.[33] We therefore have to provide some additional way of explicitly enforcing feature unification in these cases, if needed. To this end, in the following, Kallmeyer and Osswald's (2017) analysis will be slightly extended.

4.1 *Edge feature unification on final derived tree*

Edge feature unification is performed only on the final derived tree; during derivation, only node feature structures unify whenever two nodes merge because of substitution, sister adjunction or wrapping substitution. The idea is the same as that of top and bottom feature structures in LTAG (Vijay-Shanker and Joshi 1988). In LTAG, each node has a top and a bottom feature structure. If something adjoins, the two get separated. On the final derived tree, for each node, the top and bottom feature structure have to unify. This creates a means to express obligatory adjunction constraints via a mismatch between top and bottom at the respective node. In TWG, structures are flatter and we use sister adjunction. Therefore, instead of top and bottom on

[33] Edge features are used for instance to keep track of operators and periphery elements and, since substitution nodes are (usually) full projections, they should act as islands concerning these aspects.



(a) Unification between neighbouring edges

(b) Feature unification on the left fringe

(c) Feature unification on the right fringe

Figure 14: Final feature unifications

nodes, Kallmeyer and Osswald (2017) use features on the left and the right of edges in order to express constraints on sister adjunction. On the final derived tree, for two neighbouring edges, the right feature structure of the left edge and the left feature structure of the right edge have to unify. Consequently, a mismatch between features on two neighbouring edges acts as an obligatory adjunction constraint for sister adjunction.

We will use the left and right features on edges in order to express and pass the requirement to be modified by a relative clause. In the following we will notate edge features on the daughter node of the corresponding edge, embedded under features LEFT and RIGHT. This means that the final edge feature unification amounts to unifications between specific LEFT and RIGHT features on the nodes. Substitution nodes block unification of edge features (Kallmeyer and Osswald 2017) but, if the root node of the tree that is added by substitution has explicit LEFT and RIGHT features, one can nevertheless have specific features shared between lower and higher edges. (This last option is not used by Kallmeyer and Osswald 2017.) We will use this, in combination with an additional mechanism that allows nodes to look into the left/right edge features on their leftmost/rightmost daughters respectively, as a means to model the obligatory adjunction of an extraposed relative clause in (31).

Let us briefly explain how edge features work (Kallmeyer and Osswald 2017), in particular how they unify on the final derived tree (see Figure 14). As mentioned, nodes can have, as part of their feature structure, special features LEFT and RIGHT. In the final derived tree, the LEFT feature of a node *v* unifies with the RIGHT feature of its immediate sister to the left (see Figure 14a). Furthermore, the LEFT feature of a node v that does not have a sister to the left unifies with the LEFT feature of the mother of v, provided this mother is not the root node of an elementary tree or the lower node of a d-edge (see Figure 14b). Similarly, the RIGHT feature of a node v that does not have a sister to the right unifies with the RIGHT feature of the mother of v, again provided this node is not the root node of an elementary tree or the lower node of a d-edge (see Figure 14c). (These feature unifications along the left (resp. right) fringe are independent from whether the lower node has a sister to the right (resp. left), i.e., they are also performed for unary edges.) Finally, whenever we substitute a tree with root v into a substitution node v', the complete feature structures of the two unify, including the features LEFT and RIGHT. This gives us the means to share features even across substitution sites by stating this feature sharing explicitly.

We assume, slightly extending the approach of Kallmeyer and Oswald (2017), that not only do substitution nodes (which are often full projections) block automatic edge feature unification, except if stated otherwise, but so too do the daughters of root nodes in adjunct trees (for instance the CLAUSE_{peri} node below a NUC^* node in a restrictive relative clause tree adjacent to its noun antecedent), and, furthermore, so too do the daughters of root nodes in trees where the only other daughter is linked to the root by a d-edge (as is the case for extraposed relative clauses). This makes sense given that these nodes are often also full projections, for instance in the case of relative clauses, where a clause coming with its own operator projection, i.e., aspect, tense, etc. information, is added to a NUC_N node or, in the case of extraposition, to a CLAUSE node that has its own separate operator projection.

Edge features are mainly used to express obligatory or selective adjunction constraints for sister adjunction. Figure 15 shows for instance how to enforce the adjunction of a tense operator using a boolean edge feature TNS that signals the presence/absence of tense depending on whether it has a value + or −. L and R are short for LEFT and RIGHT. In the example in Figure 15, the value − of the TNS feature on the left of the NUC node unifies with the one on the right of the NP *the girl* (variable $\boxed{3}$), which also occurs on the left of this NP.[34] Since

[34] Note that TNS is not a feature of the NP node but a left/right feature of the

this is the leftmost daughter of the CORE node, the TNS value also unifies with the one to the left of CORE, i.e., embedded in the L feature of the CORE. With the adjunction of *does*, unification of the feature on the left of the CORE with the one on the right of the next sister to the left is possible, as both feature structures are [TNS −]. To its left, the tense operator *does* signals the presence of tense (TNS = +), which can unify with the feature on the right of the PrCS node (which also signals the presence of tense and thereby expresses the requirement of a tense operator). With these features, a tense operator has to adjoin somewhere between the NUC and the PrCS and there cannot be more than one tense operator. Of course, this assumes that tense operators always come with features [L [TNS +], R [TNS −]].[35]

Features LEFT and RIGHT are supposed to represent features on edges, even though they are notated on the nodes. Their unification does not interact with the proper node features, at least not automatically. But sometimes a node should be able to look into the LEFT feature of its leftmost daughter or the RIGHT feature of its rightmost daughter (in the final derived tree). To this end, in addition to the edge features from Kallmeyer and Osswald (2017), we introduce further node features LEFT-DAUGHTER-EDGE and RIGHT-DAUGHTER-EDGE (LD-EDGE and RD-EDGE for short), which are processed as other features in the context of unifications triggered by substitution or sister adjunction and for which the following holds (see Figure 16): On the final derived tree, the LD-EDGE feature of a node that has daughters unifies with the feature LEFT on the leftmost daughter and the feature RD-EDGE unifies with the feature RIGHT on the rightmost daughter.

---

edge between the CORE node and the NP node. The identity between the left and right TNS values signifies that the lower node (i.e., the NP) does not change tense in any way, which is why the information about presence/absence of tense is passed across the edge.

[35] As pointed out by a reviewer, in some cases, in particular in English, tense is contributed by more than one element. Such an example is (i).

(i) He should have enjoyed the trip.

In (i), both *should* and *have* would contribute the the overall tense of the sentence (represented by some node feature on the CORE and CLAUSE node, called for instance TENSE), while *should* would be treated as the element that satisfies the requirement of a tense operator encoded in the TNS edge feature.

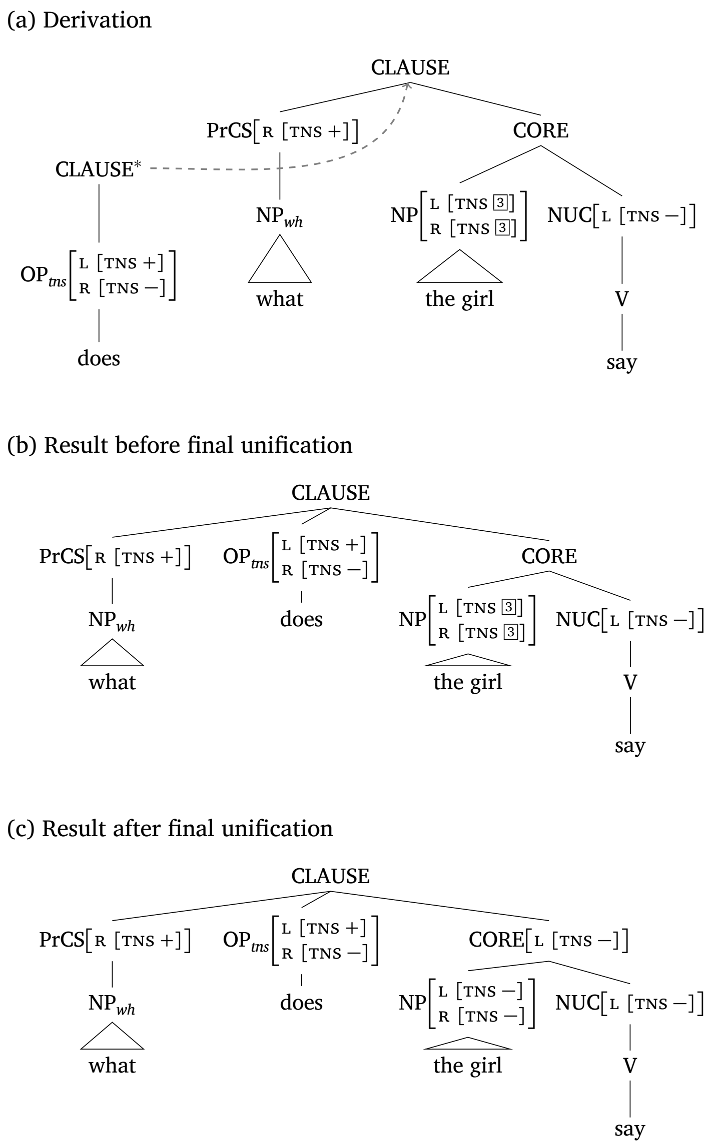


Figure 15: Obligatory adjunction of a tense operator

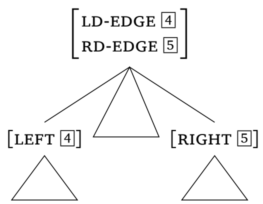


Figure 16:
Final unifications between edge and node features

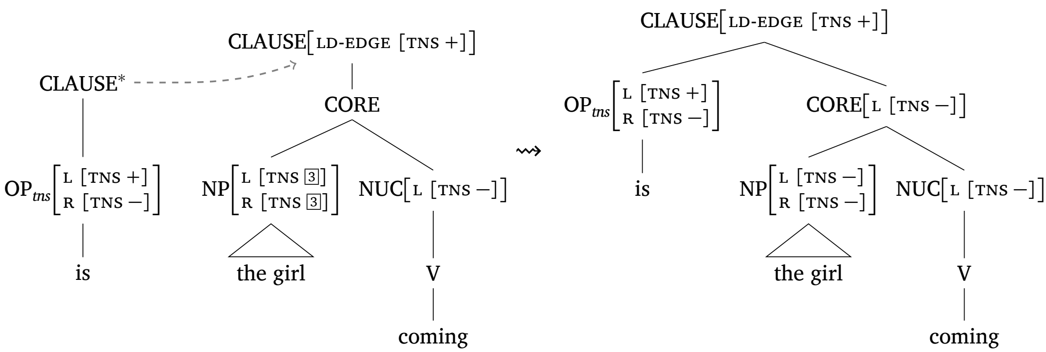


Figure 17: Obligatory adjunction of a tense operator with LD-EDGE features ("Is the girl coming?")

Note that these features are not only needed in our special case of obligatory relative clauses but also in other cases, for instance when checking for the obligatory adjunction of a tense operator in a case where the tense operator does not have a PrCS sister to the left. Such a use of LD-EDGE in order to express the requirement to have a tense operator somewhere in the tree below is given in Figure 17. Here, the tense operator adjoins to the left of the leftmost daughter of the root node. Consequently, its requirement cannot be expressed using only edge features as in Figure 15. The additional feature LD-EDGE, however, allows us to formulate constraints for the left feature structure on the leftmost edge below the root node (after derivation).[36]

[36] Note that, within LTAG, such obligatory adjunctions are handled via the top and bottom features on nodes and via the distinction between root and foot node in adjoining trees. This is why LTAG does not need features such as LD-EDGE and RD-EDGE with special unification treatments. But this is also why LTAG necessarily generates binary structures when using adjunction. See for instance XTAG Research Group (2001) for a range of analyses that model obligatory adjunction via a mismatch between top and bottom feature structures.

Other cases where a single element has to be adjoined exactly once or at most once below the root node are for instance clause linkage markers (CLM) such as *to* in (32a) and *that* in (32b).

(32) a. He promised to come.
b. He promised that he would come.

4.2 *An analysis of extraposed obligatory relative clauses using edge features*

In order to capture the requirement for obligatory relative clauses, we introduce a binary feature that expresses that a relative clause has been found or has to be found, REL-CL-EXISTS or REL-EX for short (values + or −). Using the above-mentioned features L(EFT) and R(IGHT) and their unifications on the final derived tree, the relevant constraints can be captured as follows: An NP that requires an extraposed relative clause carries features L(EFT) [REL-EX +] and R(IGHT) [REL-EX −], thereby indirectly expressing that somewhere to the right a relative clause has to be found. The requirement LEFT [REL-EX +], stating that there is (or, rather, has to be) an extraposed relative clause in the final derived tree, is passed upwards to the left while the lack of a relative clause so far, RIGHT [REL-EX −], is passed to the right and upwards. We put the latter on the edge between NUC_N and N (notated, as mentioned, on the N node), which means that it gets passed upwards only if no relative clause adjoins to NUC_N (as in (31a), where the NP node would have a R [REL-EX +] feature, i.e., there would not be any requirement for an extraposed relative clause). An example is the left tree in Figure 18, where the final edge feature unification leads to $\boxed{3} = -$. For other roots of NP trees, these features are not specified, leaving it open whether a relative clause is added.

The tree into which the NP substitutes makes sure any REL-EX features on edges get percolated via edge feature unification across non-leaves towards the outermost nodes and then upwards (in our example, because the NP is the leftmost daughter of the CORE node, this is given anyway for the LEFT feature on the NP, and for the

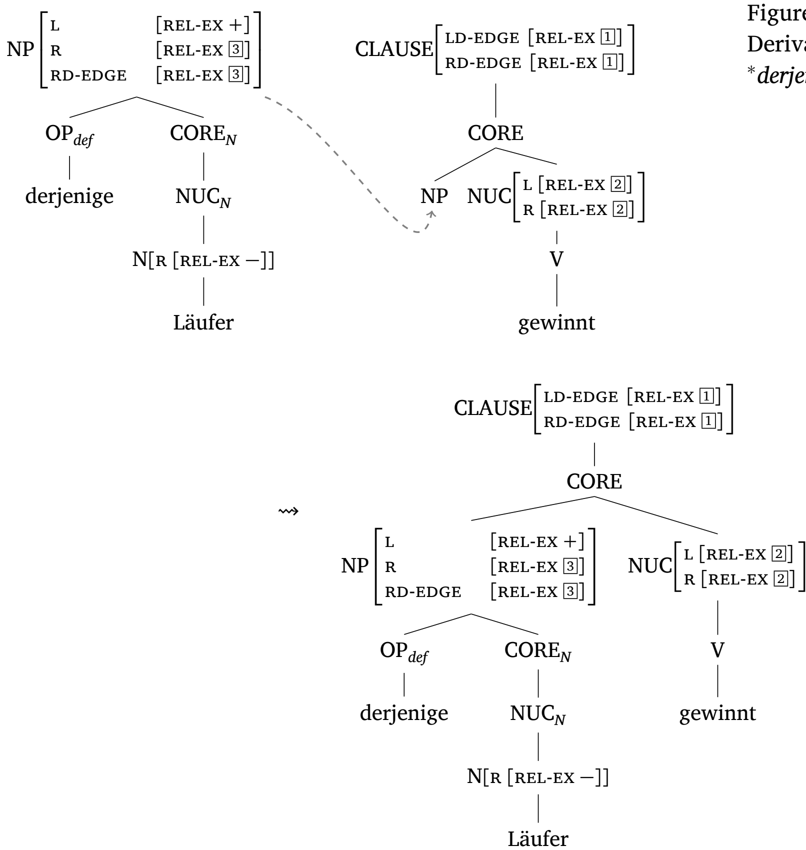


Figure 18: Derivation of (31c) **derjenige Läufer gewinnt*

RIGHT feature, we have to make sure it can percolate[37] via the NUC node, which is done in the LEFT and RIGHT features on that node). At the root, unification of the information coming from the left (the requirement) and coming from the right (the information on existing relative clauses) is then unified. This last unification, which matches the requirement with what has been found, is done by stating on the CLAUSE node that the REL-EX value on the left of the edge to the leftmost daughter (feature LD-EDGE) has to unify the REL-EX value on the right of the edge to the rightmost daughter (feature RD-EDGE).

[37] Strictly speaking, there is no percolation here but only feature unification. The term "percolate" is used to indicate that a specific value is specified in one place and, due to unification, gets passed to other places.

The trees for NP and matrix verb in our example are given in Figure 18, which sketches the derivation for the ungrammatical (31c). If we perform the substitution and then end the derivation, the final unification on the derived tree will fail for two reasons: firstly, the LEFT feature on the NP node unifies with the LEFT feature on its mother and the LD-EDGE feature of the CLAUSE node (as a result, we obtain $\boxed{1} = +$); and, secondly, the RD-EDGE feature on the CLAUSE node (now REL-EX +) has to unify with the right edge feature on the rightmost daughter, which would be the RIGHT feature on the CORE and the NUC nodes (REL-EX −). Consequently, adding a further daughter of the CLAUSE node to the right of the CORE is obligatory, in order to change the REL-EX value on the rightmost daughter of the CLAUSE node to +.

The derivation of (31b) with the extraposed relative clause is given in Figure 19 (the previously introduced features NUC-ID, N-ID and PERI-SCOPE are left aside here for the sake of readability). Instead of combining the NP *derjenige Läufer* directly with the NP substitution slot in the *gewinnt* tree, we have to substitute it into the NP leaf in the relative clause tree, which is then in turn substituted into the subject slot of *gewinnt* via wrapping substitution, adding at the same time the relative clause to the root. The $\text{CLAUSE}_{\text{peri}}$ node has a feature R [REL-EX +], which signals that a relative clause has been attached.

Figure 20 shows the final derived tree: Figure 20a gives the features before final unification, while Figure 20b specifies them after the final edge feature unification.

Let us briefly inspect the structure of the NPs of type *derjenige* (a pronoun) and *derjenige N* (a full NP) more closely. The two cases are given in Figure 21 and Figure 22. In the pronoun case (Figure 21), we can put the + and − values on the left and right of the NP root directly on that node. In the case of *derjenige N*, in the tree for *derjenige*, which adjoins at the NP root of the nominal tree (see Figure 22a), we can put the value + for REL-EX on the left of the OP_{def} node and − on the right. In the final feature unification, once the derivation is finished, this latter gets passed down on the left of the CORE_N – NUC_N – N spine. On the N node we pass it explicitly to the right (see variable $\boxed{8}$ in Figures 22a and 22b). From there, if nothing intervenes (for example a relative clause attaching at NUC_N), the value − gets passed upwards and ends up in the right of the root NP. The NP root

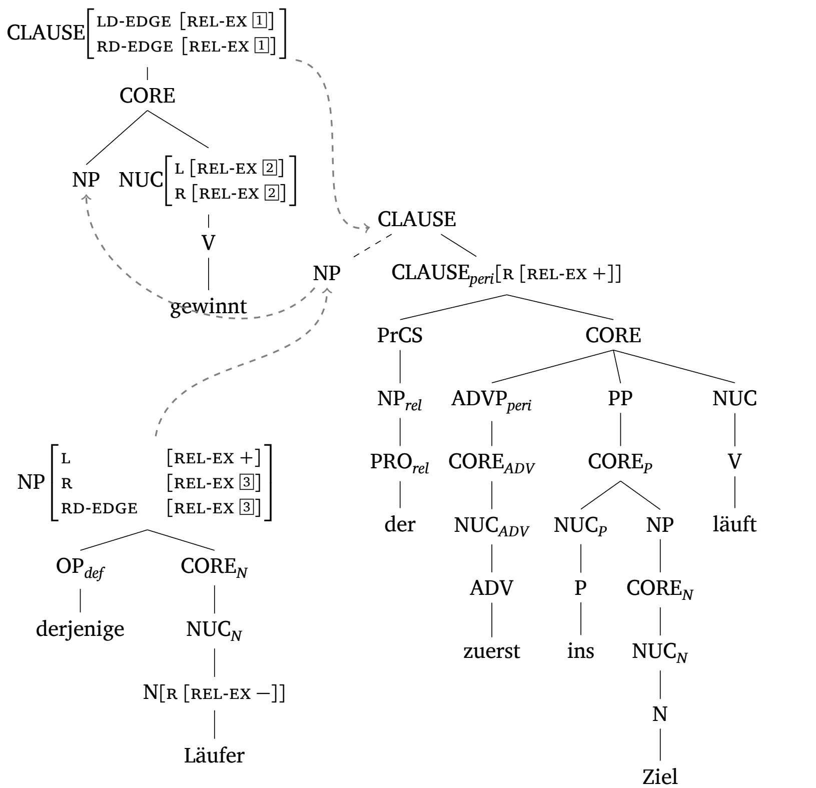
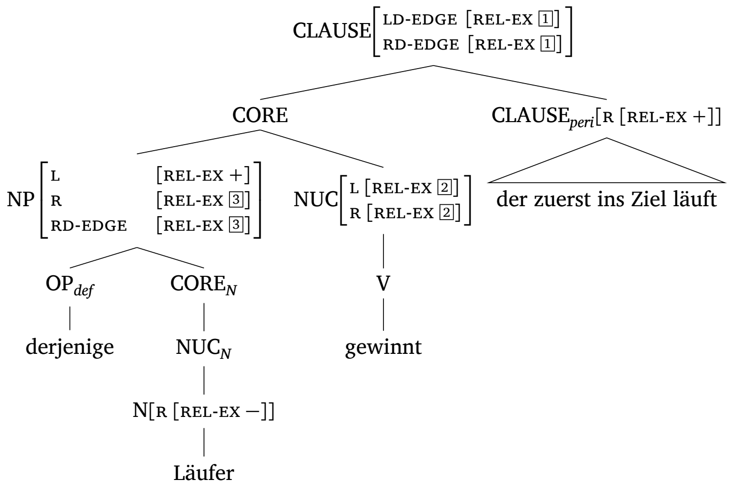


Figure 19: Derivation for (31b)

node in the nominal tree does not fix the left and right values; it just states that the one coming from the leftmost daughter has to unify with the one under LEFT on that node and, respectively, the one coming from the rightmost daughter has to unify with the one under RIGHT on the NP node (see variables 6 and 7 in Figures 22a and 22b). As a consequence, if there is no requirement, the two will be equal while in the case of a *derjenige* operator adjoining, the left will be + and the right − (see the derived tree after feature unification in Figure 22c).

Now let us go back to the overall way to model requirements for extraposed relative clauses with edge features. There is still something missing with the analysis proposed so far: It only guarantees that whenever we have an NP of the form *derjenige (N)*, we will also have a relative clause. But if this relative clause is extraposed, it does however not guarantee that it is a modifier of the NP in question, it can also

(a) Before final edge feature unifications:



(b) After final edge feature unification:

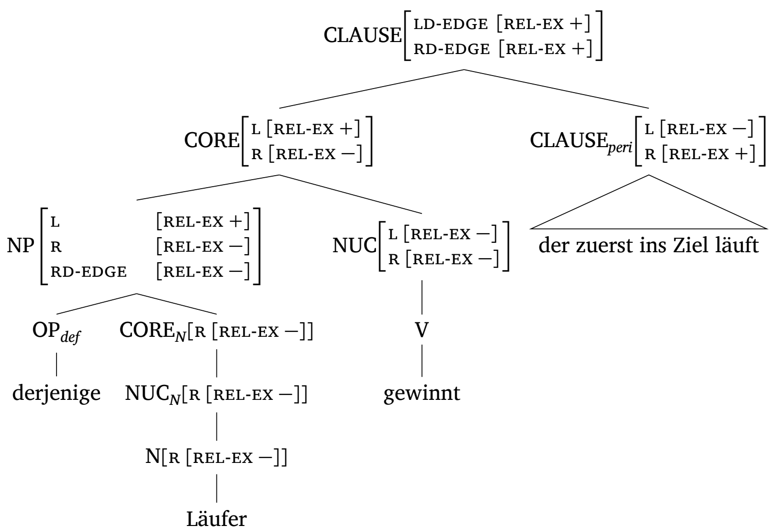


Figure 20: Resulting derived tree for (31b)

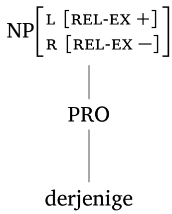


Figure 21: Internal structure of NP of the form *derjenige*

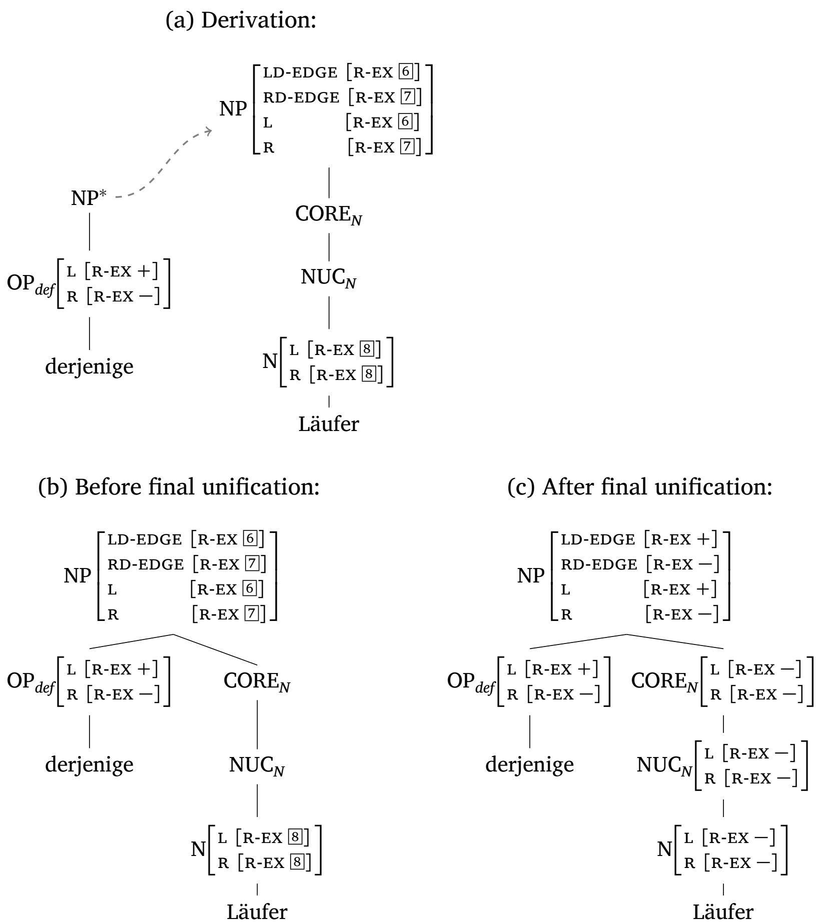


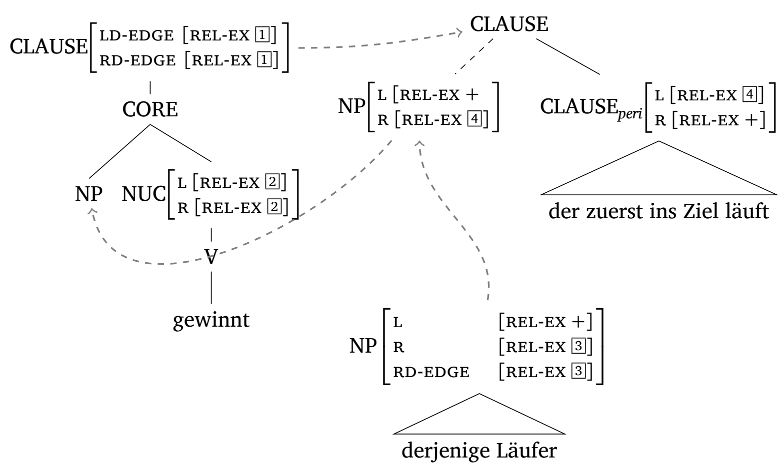
Figure 22: Internal structure of NPs *derjenige N* (R-EX is short for REL-EX): derivation and results before and after final unification

modify another NP. In other words, the NP leaf coming with the elementary tree of the relative clause does not necessarily merge with the one of the corresponding antecedent NP. We would for instance also be able to derive the ungrammatical (33) (our own example), where the agreement features of the relative clause do not match those of the NP *demjenigen Mädchen* ('the one girl').

(33) *Der Junge gibt demjenigen *Mädchen* ein Buch, [*der*
the boy.M gives the.one.N girl.N a book.N who.M
zuerst den Raum betritt].
at.first the room enters.

In order to enforce a substitution of the correct antecedent NP into the NP node of the relative clause, we add an identity requirement for the REL-EX feature on the left of the edge to the CLAUSE$_{peri}$ node and the REL-EX feature on the right of the antecedent NP. In the case of an extraposed obligatory relative clause, the value on the left of the edge to the CLAUSE$_{peri}$ node is −, consequently, the NP also has to have the right REL-EX value −. In addition, we impose that the left REL-EX value on the NP node is +. Figure 23 shows this extension for our previous example (31b). If we have an NP of the type *derjenige N* in the sentence, only this NP will have different features REL-EX to its left (value +) and its right (value −), whereas all other NPs have equal values and thereby just pass along what they see to their right/left

Figure 23: Enforcing substitution of correct antecedent NP for extraposed obligatory relative clauses



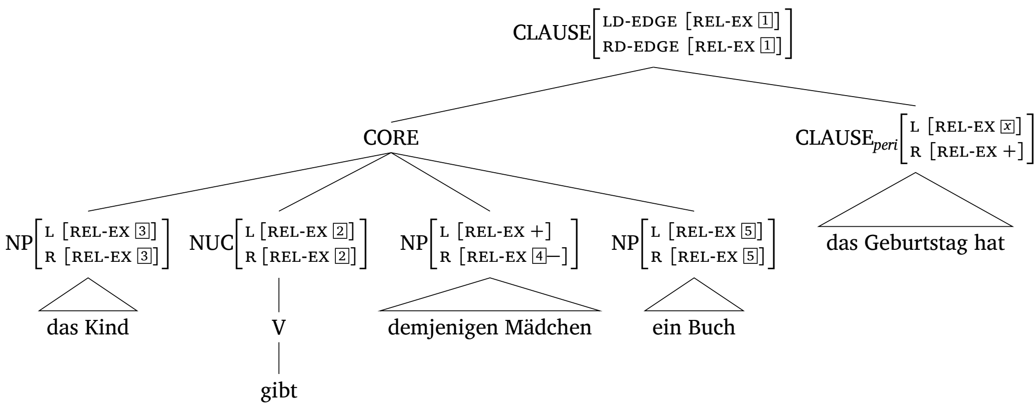


Figure 24: Derived tree for (34) before final feature unifications

respectively. If there is no such NP in the sentence, all REL-EX feature values will be the same, namely either undefined (if no extraposed relative clause is added) or + (if one is added).

As a further example, let us have a look at the tree we would derive for (34), where we have three NPs, all of them with the same agreement features (GEN = *n*, NUM = *sg*) and therefore in principle all of them possible antecedents for the extraposed relative clause (we left aside agreement features for reasons of readability but of course they are taken into account when choosing the antecedent NP, see Section 3.4.3). But since one of the NPs is an NP of the type *derjenige N*, this one necessarily has to become the antecedent.

(34) Das Kind gibt demjenigen *Mädchen* ein Buch, [*das*
the child.N gives the.one.N girl.N a book.N who.N
Geburtstag hat].
birthday has
'The child gives a book to the girl whose birthday it is.'

Figure 24 shows the derived tree before final feature unifications. More precisely, it represents the three different derived trees we would obtain depending on which of the argument NPs was substituted into the NP slot of the extraposed relative clause, i.e., which of the three NPs was chosen as antecedent: The variable x in this tree is a placeholder for either 3 or 4 or 5 depending on whether the *Kind* NP, the *Mädchen* NP or the *Buch* NP was substituted into the relative clause

NP antecedent slot. We will see that the final feature unifications will exclude the first and the last possibility.

Concerning final feature unifications, no matter which NP has been targeted, we always obtain $\boxed{4} = \boxed{5} = \boxed{x} = -$ and $\boxed{2} = \boxed{3} = \boxed{1} = +$ because of the outwards and upwards percolation starting from the NP node of *demjenigen Mädchen*. Consequently, since the NP node in the relative clause elementary tree states that its REL-EX value under R has to unify with the REL-EX value under L at the CLAUSE$_{\text{peri}}$ node, and since the latter necessarily is −, the *demjenigen Mädchen* NP is the only possible antecedent. All other NPs have identical REL-EX values on their left and their right due to the internal structure of the NP. They could be antecedents of non-obligatory relative clauses, with the two REL-EX features in question having a value +.

In clauses where an extraposed relative clause is present but is not required because none of the NPs is a *derjenige N* NP, all the REL-EX values would become + since, starting from the right feature on the CLAUSE$_{peri}$ they would be passed around. On the other hand, in clauses with no extraposed relative clause (and no requirement for adding one), the REL-EX values would all unify but remain unspecified.

A potential problem of this approach might however be that the feature REL-EX only expresses the requirement for an extraposed relative clause and whether the requirement has been met so far. It does not specify which NP has triggered the requirement. Therefore, this approach hypothesizes that we have at most one NP in a CLAUSE that requires an extraposed relative clause.

The examples in (35) (our own constructed examples) suggest that we can have more than one NP of the form *derjenige (N)* with corresponding extraposed relative clauses in a single sentence but not in the same clause. In the examples in (35), the agreement features of the pronouns *derjenige* and *derjenigen* and of the two relative pronouns leave only one option for the choice of antecedent NPs for the two relative clauses, namely the one expressed by the coindexations. The two examples (35a) and (35b) with the antecedent NPs being arguments of the same verbs are both ungrammatical. It seems that we can have more than one such NP in a sentence only if these NPs (and their corresponding extraposed relative clauses) occur in different CLAUSE subtrees, as in (35c) and (35d), which is possible with our analysis.

(35) a. **Derjenige*$_1$ schenkt *derjenigen*$_2$ ein Buch, [*die als erstes den*
the.one.M offers the.one.F a book who.F as first the
Raum betritt]$_2$, [*der die Wette verloren hat*]$_1$.
room enters who.M the bet lost has
'The one who lost the bet offers a book to the one who enters the room first.'

b. **Derjenige*$_1$ schenkt *derjenigen*$_2$ ein Buch, [*der die Wette verloren hat*]$_1$, [*die als erstes den Raum betritt*]$_2$.

c. Hans, der heute *denjenigen* abholt, [*den er gestern*
Hans who today the.one.M fetched who.M he yesterday
angerufen hat], schenkt nächste Woche *derjenigen* ein Buch,
called has offers next week the.one.F a book
[*die die Wette gewonnen hat*].
who.F the bet won has
'Hans who fetched today the one whom he called yesterday will next week offer a book to the one who won the bet.'

d. Hans schenkt *demjenigen*$_1$ ein Buch, [*der der Bruder von*
Hans offers the.one.M a book who.M the brother of
derjenigen$_2$ *ist*, [*die die Wette gewonnen hat*]$_2$]$_1$.
the.one.F is who.F the bet won has
'Hans offers a book to the one who is the brother of the one who has won the bet.'

Note, however, that it is hard to tell whether more than one extraposed obligatory relative clause attaching to the same clause should be possible, based only on the examples in (35). It might be that this restriction, that comes with our analysis, is a problem. Consider for instance (36) (our own example), where we have one *derjenige (N)* NP embedded in another *derjenige (N)* NP, both with extraposed relative clauses that attach to the same CLAUSE node. Example (36) seems more acceptable than (35a) and (35b).

(36) Winston hat dasjenige *Buch*$_1$ von demjenigen *Autor*$_2$
Winston has the.one.N book.N of the.one.M author.M
ausgeliehen, [*der eigentlich verboten ist*]$_2$, [*das er aber*
borrowed who.M actually forbidden is that.N he but
überraschenderweise in der Bibliothek entdeckt hatte]$_1$.
surprisingly in the library discovered had

'Winston has borrowed the one book of that author who is actually forbidden, but which he had surprisingly discovered in the library.'

The fact that we needed some non-local feature sharing (via edge feature unification) here in order to capture the obligatoriness of certain relative clauses might seem contradictory to our initial claim that with RRG/TWG, the relation between a relative clause and its antecedent NP can be captured locally, within one elementary tree. Our analysis, however, still captures this relation locally; the only aspect that the shared features capture is the request for some extraposed relative clause. It does not indicate the exact antecedent that requires that relative clause. In this respect, the RRG analysis proposed here still differs fundamentally from HPSG analyses, as in Walker (2017), where information about the actual antecedent NPs is percolated (see Section 5.2 below). In our case, the percolated feature is only a single binary feature, while in the HPSG analyses, it is a list-valued feature that can, in principle, have arbitrarily many different values.

4.3 *An alternative local feature-based analysis*

The reason why we introduced the feature percolation mechanism for REL-EX in the preceding section was that the lower NP node of the relative clause cannot directly change the value of that feature from − to + at the NP root node of the *derjenige*-NP. This is not possible simply because the feature structures of the NP-node of the *derjenige*-NP, those of the relative clause antecedent NP slot and those of the NP argument slot in the tree of the matrix verb all unify. Unification is monotonic, i.e., it can only add information. Changing features is only possible from a node to a different node or between edge features of for instance sister nodes.

Given that the relative clause and its antecedent NP are, however, linked at the NP antecedent node via substitution, it would be more in line with the overall ideas of the grammar theory to take care of the relative clause requirement in some local way at that node.

Furthermore, we have seen that the feature percolation approach to extraposed obligatory relative clauses comes with the constraint to

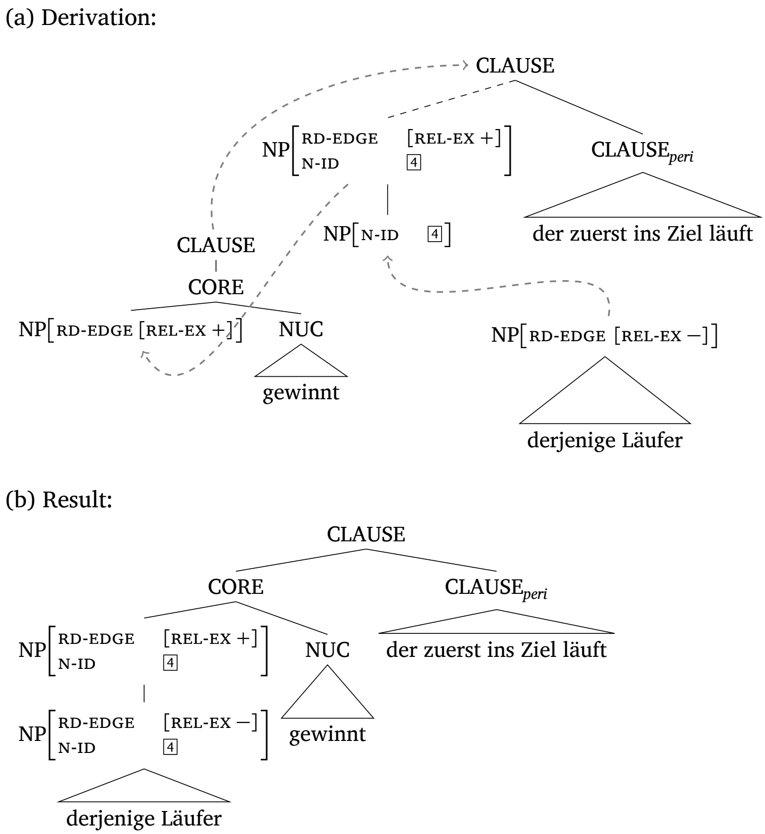


Figure 25: Local feature use for enforcing extraposed relative clause

have at most one such obliatory extraposed relative clause attaching at the same CLAUSE node, which might be too restrictive.

This observation leads to a different, more local, way of using the REL-EX feature in order to enforce adding a relative clause, exemplified in Figure 25. We replace the single NP antecedent slot in the tree of the extraposed relative clause with two NP nodes, one a daughter of the other, and the higher one carries a feature RD-EDGE [REL-EX +]. That is, no matter whether the NP tree we insert below has a request or not for a relative clause, the root NP node of the tree that fills the argument slot will have RD-EDGE [REL-EX +]. This is also what we require for every NP argument slot, which means that a tree such as the one for *derjenige Läufer* in Figure 25 cannot be substituted into such a slot. For all other features, the two NP-nodes in the relative clause tree require identity, except edge features, i.e., features LEFT and RIGHT. In order to signal that they are thought of as two copies of the same node, different only in terms of edge features, we give them the same node identifier (see the shared feature N-ID).

Note that the semantics of the feature REL-EX is slightly different with this approach, REL-EX $= -$ now signifies that there is a so far unsatisfied request for a relative clause, while REL-EX $= +$ means that all requests for relative clauses below an NP node have been satisfied. The latter holds also in cases where there are no requests. That is, we can interprete REL-EX as meaning something like “any request for an extraposed obligatory relative clause satisfied?”.

This is a simple local way to enforce adding an extraposed relative clause in the case of a *derjenige* NP that does not yet contain a relative clause. The inconvenience is that we have added an extra NP node and a unary branch to the tree. However, if we assume that features LEFT, RIGHT, LD-EDGE and RD-EDGE can actually be ignored and therefore deleted once the derivation including the final edge feature unifications is finished, we could perform a merging of identical nodes linked by a unary immediate dominance edge in the derived tree.[38]

This analysis still allows for antecedent NPs with multiple extraposed relative clauses as exemplified in (14). In these cases, we would obtain three NP nodes in a unary spine in the derived tree that would collapse into one node after final feature unifications.

So far, this analysis does not restrict the number of obligatory relative clauses that can attach below the same clause to one. For such a constraint, we could use a simple boolean edge feature on the edge from CLAUSE to CLAUSE_{peri}. However, it is not clear whether this constraint really holds (see the discussion of example (36) above). It might

[38] Note that this local analysis is close to what is performed in the RRG parser implementation described in Bladier et al. (2020b), where such unary branches with copies of nodes are created solely for a technical, parser-internal reason, namely because the parser does not allow d-daughters (i.e., the lower nodes of d-edges) to be at the same time substitution nodes. That is, the parser in Bladier *et al.* (2020b) introduces a temporary daughter (with an identical label) in these cases that gets deleted after parsing.

actually be an advantage of this approach, compared to the previous, feature percolation based one, that several extraposed obligatory relative clauses attaching to the same CLAUSE node are possible.

It is hard to tell which solution is better: the feature percolation solution in Figure 23 or the one with the extra NP node in the antecedent part of the relative clause trees (see Figure 25). So far, the TWG formalization of RRG is inspired by the idea that long-distance dependencies should arise from tree wrapping (and not from unbounded feature percolation). This points towards the latter option, even though it comes with a slightly unusual unary branch. Concerning predictions that the two approaches make, it might be an advantage of the second that it does not exclude more than one obligatory relative clause at the same CLAUSE node. This, together with a preference for local solutions leads us to opting for this latter solution, keeping in mind that in the final derived tree, we can merge identical nodes linked by a unary branch.

COMPARISON TO OTHER APPROACHES 5

Lexicalized Tree Adjoining Grammar (LTAG) 5.1

An LTAG (Joshi *et al.* 1975; Joshi and Schabes 1997) is also a tree rewriting grammar, as TWG, but with different composition operations. Trees can be combined either via *substitution* or via *adjunction*. The latter consists of replacing an internal node with an *auxiliary tree*, which is a tree with a non-terminal leaf node marked as *foot node*. When adjoining, the subtree below the adjunction site ends up below the foot node. Adjunction is more powerful than sister adjunction. It serves, roughly, two purposes: on the one hand, it is used to add modifiers and functional operators; on the other hand, it realizes long-distance dependencies by adding material in between two nodes that come from the same elementary tree. In the case of RRG-TWG, the former is modelled with sister adjunction and the latter with wrapping substitution. Note that the tree added in an adjunction is a (possibly derived) auxiliary tree, i.e., a tree with a single foot node. Its root and foot node always originate from the same elementary tree.

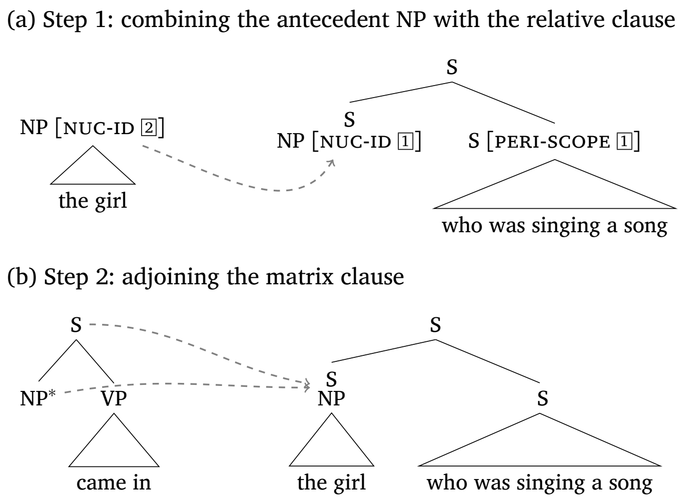


Figure 26: Sketch of an LTAG adjunction analysis for extraposed relative clauses along the third TWG analysis above

To our knowledge, there is no LTAG analysis of extraposed relative clauses. Papers that deal with relative clauses in the context of LTAG are concerned with cases where the relative clause has a complex internal structure, a long-distance dependency for example or a relative pronoun that is embedded into a complex relative phrase (Kahane 2000; Han 2002; Kallmeyer 2003).

LTAG does not easily provide an analysis for extraposed relative clauses that combines antecedent NP or antecedent noun and relative clause in one elementary tree. It is too restricted to provide such a solution, at least with standard LTAG trees: An analysis along the lines of our TWG analysis above (the third analysis, see Sec. 3.3) would amount to adjoining the matrix clause into the relative clause tree, thereby separating the antecedent NP slot from the relative clause. This possibility is sketched in Figure 26. The relative clause tree has a substitution node for the antecedent NP. Deviating slightly from standard TAG, one could allow different top and bottom categories (here: S and NP), which can be seen as different CAT features in the two feature structures. When substituting the antecedent NP into the relative clause tree (Figure 26a), the root feature structure (CAT = *np*) of the incoming *the girl* tree unifies with the lower feature structure, which leads then to the tree on the lower right in the second derivation step. In the second step (Figure 26b), instead of wrapping this around the *came in*

(a) Step 1: combining the antecedent NP with the matrix clause

(b) Step 2: adjoining the relative clause

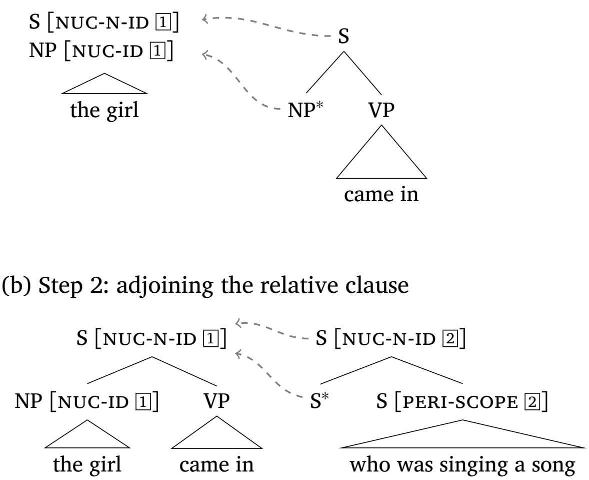


Figure 27: Sketch of an LTAG adjunction analysis for extraposed relative clauses along the second TWG analysis above

tree, as in TWG, one could adjoin the latter to the relative clause tree, following LTAG's general strategy of doing extraction by adjunction.

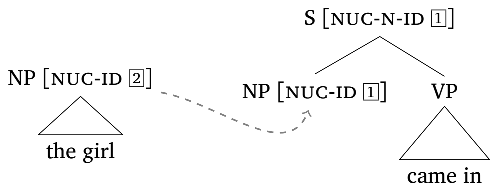
Such a solution, however, would exclude cases where the antecedent NP (the foot node in the adjoining tree) is not part of the argument structure of the matrix clause. Put differently, the root and the foot node of the adjoining tree have to come from the same elementary tree. This is a crucial difference to TWG where, in a wrapping step, the two target nodes can come from different elementary trees.

An LTAG analysis along the lines of our second analysis option (Section 3.2) would, roughly, look as exemplified in Figure 27. Here, again, the S that is the adjunction site for the relative clause and the NP antecedent node have to be part of the same elementary tree.

Yet another possibility would be to let the matrix verb anticipate the adjunction of an extraposed relative clause for one of its argument NPs, see Figure 28.

In all three cases, it would be possible to further embed the antecedent NP but only by adjoining material in between the S and the NP node, not by substitution. A second problem is that the matrix clause tree has an NP leaf for the antecedent NP, which, according to standard LTAG principles, means that this NP is an argument. But

(a) Step 1: combining the antecedent NP with the matrix clause



(b) Step 2 (adjoining the relative clause): see step 2 in Figure 27

Figure 28: Sketch of an LTAG adjunction analysis for extraposed relative clauses where the matrix verb anticipates the extraposed relative clause

the antecedent NP can also be part of an adjunct, in which case it is unclear how to model that.

A more severe problem of all three options would be that they exclude multiple extraposed relative clauses with different antecedent NPs since LTAG elementary trees can have at most one foot node, i.e., we cannot have more than one NP* node in a tree (this limits the options in Figures 26 and 27), and we can provide at most one value for the NUC-N-ID feature at the S node of the matrix verb (this limits the options in Figures 27 and 28). This concerns the second crucial difference to TWG where we can have more than one d-edge stretching across a node.

The only solution TAG can easily offer is the anaphoric approach, where a subsequent process on the derived tree determines the antecedent of the relative clause. This would look like step 2 in Figure 27 but without the features that link the antecedent NP to the scope of the relative clause. It is, however, not clear how such a subsequent process of relating relative clauses with their antecedents could look, given the limited possibilities coming with LTAG's use of feature structures.

5.2 *Head-Driven Phrase Structure Grammar (HPSG)*

Following Kiss' 2005 theory of Generalized Modifiers in HPSG, Walker (2017) proposes an HPSG analysis along the following lines: The "extraposed relative clause is base-generated", and "an anchor that percolates throughout the tree is used to establish the relationship between the relative pronoun and its antecedent" (page 159). A set-valued attribute ANCHORS is used to collect referential phrases that are antecedents of relative clauses. Like HPSG's SLASH feature, it is part of the values of the attributes INHERITED and TO-BIND under NONLOCAL. The anchors are passed upwards as elements of INHERITED ANCHORS, and when encountering a relative clause, an appropriate element on the anchors list is identified with an index value on the ANCHORS set under the feature MOD of the relative clause.

In order to account for obligatory relative clauses with *derjenige (N)* NPs, Walker (2017) imposes that at the root of the entire tree, the ANCHORS set must be empty. NPs can introduce anchors but need not do so, except for *derjenige (N)* NPs where the introduction of an anchor is obligatory.

A crucial difference to the RRG-TWG approach proposed in this paper is that TWG makes use of its extended domain of locality, connected to the operation of wrapping substitution, in order to group the antecedent NP node and the relative clause into one elementary tree. This would not be possible for HPSG, which is lacking an extended domain of locality. On the other hand, a SLASH or ANCHORS feature percolation analysis along the lines of HPSG is not possible for RRG-TWG because of the more restricted types of feature structures used on nodes and edges. TWG uses only a finite set of feature structures, which is crucial for not extending its generative capacity beyond mildly context-sensitive languages. This, however, excludes set- or list-valued features. Even the percolation techniques proposed above in our first approach for dealing with extraposed obligatory relative clauses assume that there is at most one such request or pronoun that is dealt with in a specific node. (There might be of course more than one in an entire tree but in different parts of the tree.)

This illustrates the fundamental difference between, on the one hand, tree rewriting formalisms that come with an extended domain of locality (TAG, TWG) but with restricted tree composition operations and therefore a restricted generative capacity; and, on the other hand, formalisms such as HPSG without a notion of extended domain of locality but with an increased generative capacity due to a highly expressive logic. The former frequently enable a local analysis of nonlocal dependencies but are sometimes too restricted. We claim that the expressive power of TWG is sufficent to deal with a large range of phenomena in an appropriate way.

6 CONCLUSION

In this paper, we have developed an analysis of extraposed relative clauses that establishes the link between a relative clause and its antecedent NP in a local way in the sense of placing them in the same elementary building block. The analysis is formulated in the theory of Role and Reference Grammar, assuming its formalization as a Tree Wrapping Grammar. It can account for embedded antecedent NPs, multiple extraposed relative clauses, and extraposed obligatory relative clauses. We have shown that tree wrapping allows us to deal with this phenomenon in a local way, i.e., by comprising the relative clause and the slot for its antecedent NP in the same elementary tree. There is no need for unlimited feature percolation across the derived tree, even for obligatory relative clauses (if a slightly unusual form for the slot of the antecedent NP is used).

The paper contributes a detailed and formally precise analysis of extraposed relative clauses within RRG, a topic that has not been considered so far within this grammar theory. Furthermore, and even more importantly, it proposes an analysis of this phenomenon in a tree rewriting grammar formalism inspired by LTAG but extending it. It addresses the fact that a restricted tree rewriting operation such as LTAG's adjunction allows for an elegant analysis of certain long-distance dependencies (Kroch and Joshi 1987) while being in some cases too restricted. The use of tree wrapping instead of adjunction gives us a less restricted operation for long-distance dependencies that can also model rather non-local phenomena such as extraposed relative clauses in a local way, i.e., with the long-distance dependency originating from a single elementary tree.

ACKNOWLEDGEMENTS

This paper has benefitted immensely from numerous discussions with Rainer Osswald and Robert Van Valin. Their valuable comments helped considerably towards improving the analyses and the paper. Furthermore, special thanks go to three anonymous reviewers, whose detailed comments contributed a lot to better organizing the paper, including more data and being clearer and more precise in the analyses and the comparisons with other approaches. This work was carried out as a part of the research project TreeGraSP funded by a Consolidator Grant of the European Research Council (ERC).

REFERENCES

Anne ABEILLÉ and Owen RAMBOW (2000), Tree Adjoining Grammar: An overview, in Anne ABEILLÉ and Owen RAMBOW, editors, *Tree Adjoining Grammars: Formalisms, Linguistic Analysis and Processing*, pp. 1–68, CSLI.

Artemis ALEXIADOU, Paul LAW, André MEINUNGER, and Chris WILDER (2000), Introduction, in Artemis ALEXIADOU, Paul LAW, André MEINUNGER, and Chris WILDER, editors, *The syntax of relative clauses*, volume 32 of *Linguistics Today*, pp. 1–51, John Benjamins, Amsterdam and Philadelphia.

Mark R. BALTIN (2006), Extraposition, in Martin EVERAERT and Henk VAN RIEMSDIJK, editors, *The Blackwell Companion to syntax, volume 2*, pp. 237–271, Blackwell Publishing, Malden, MA.

Mark Reuben BALTIN (1978), *Toward a theory of movement rules*, Ph.D. thesis, Massachusetts Institute of Technology.

Tatiana BLADIER, Laura KALLMEYER, Rainer OSSWALD, and Jakub WASZCZUK (2020a), Automatic extraction of Tree-Wrapping Grammars for multiple languages, in *Proceedings of the 19th International Workshop on Treebanks and Linguistic Theories*, pp. 55–61, Association for Computational Linguistics, Düsseldorf, Germany, doi:10.18653/v1/2020.tlt-1.5, `https://www.aclweb.org/anthology/2020.tlt-1.5`.

Tatiana BLADIER, Andreas VAN CRANENBURGH, Kilian EVANG, Laura KALLMEYER, Robin MÖLLEMANN, and Rainer OSSWALD (2018), RRGbank: A Role and Reference Grammar corpus of syntactic structures extracted from the Penn Treebank, in *Proceedings of Treebanks and Linguistic Theories*, pp. 5–16, `http://www.ep.liu.se/ecp/155/003/ecp18155003.pdf`.

Tatiana BLADIER, Jakub WASZCZUK, and Laura KALLMEYER (2020b), Statistical parsing of Tree Wrapping Grammars, in *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 6759–6766, International Committee on Computational Linguistics, Barcelona, Spain (Online), doi:10.18653/v1/2020.coling-main.595, `https://www.aclweb.org/anthology/2020.coling-main.595`.

Daniel BÜRING and Katharina HARTMANN (1997), Doing the right thing, *The Linguistic Review*, 14(1):1–42.

Berthold CRYSMANN (2013), On the locality of complement clause and relative clause extraposition, in Gert WEBELHUTH, Manfred SAILER, and Heike WALKER, editors, *Rightward movement in a comparative perspective*, Linguistik Aktuell/Linguistics Today, pp. 369–396, John Benjamins.

Mark DE VRIES (2002), *The syntax of relativization*, Ph.D. thesis, Universiteit van Amsterdam.

Tomaž ERJAVEC, Ana-Maria BARBU, Ivan DERZHANSKI, Ludmila DIMITROVA, Radovan GARABÍK, Nancy IDE, Heiki-Jaan KAALEP, Natalia KOTSYBA, Cvetana KRSTEV, Csaba ORAVECZ, Vladimír PETKEVIČ, Greg PRIEST-DORMAN, Behrang QASEMIZADEH, Adam RADZISZEWSKI, Kiril SIMOV, Dan TUFIŞ, and Katerina ZDRAVKOVA (2010), MULTEXT-East "1984" annotated corpus 4.0, Slovenian language resource repository CLARIN.SI.

Janet D. FODOR (1978), Parsing strategies and constraints on transformations, *Linguistic Inquiry*, 9(3):427–473.

Chung-hye HAN (2002), Compositional semantics for relative clauses in Lexicalized Tree Adjoining Grammars, in *Proceedings of the Sixth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+6)*, pp. 1–10, Association for Computational Linguistics, Universitá di Venezia, `https://www.aclweb.org/anthology/W02-2201`.

Anke HOLLER (2013), D- und w-Relativsätze, in Jörg MEIBAUER, Markus STEINBACH, and Hans ALTMANN, editors, *Satztypen des Deutschen*, De Gruyter Lexikon, pp. 266–300, De Gruyter, Berlin.

Aravind K. JOSHI, Leon S. LEVY, and Masako TAKAHASHI (1975), Tree Adjunct Grammars, *Journal of Computer and System Science*, 10:136–163.

Aravind K. JOSHI and Yves SCHABES (1997), Tree-Adjoning Grammars, in Grzegorz ROZENBERG and Arto SALOMAA, editors, *Handbook of formal languages*, pp. 69–123, Springer, Berlin.

Sylvain KAHANE (2000), How to solve some failures of LTAG, in *Proceedings of the Fifth International Workshop on Tree Adjoining Grammar and Related Frameworks (TAG+5)*, pp. 123–128, Université Paris 7, `https://aclanthology.org/W00-2017`.

Laura KALLMEYER (2003), LTAG semantics for relative clauses, in Harry BUNT, Ielka VAN DER SLUIS, and Roser MORANTE, editors, *Proceedings of the Fifth International Workshop on Computational Semantics IWCS-5*, pp. 195–210, Tilburg.

Laura KALLMEYER (2016), On the mild context-sensitivity of *k*-Tree Wrapping Grammar, in Annie FORET, Glyn MORRILL, Reinhard MUSKENS, Rainer OSSWALD, and Sylvain POGODALLA, editors, *Formal Grammar. 20th and 21st*

International Conferences, FG 2015, Barcelona, Spain, August 2015, Revised Selected Papers. FG 2016, Bozen, Italy, August 2016, Proceedings, volume 9804 of *Lecture Notes in Computer Science*, pp. 77–93, Springer.

Laura KALLMEYER and Rainer OSSWALD (2017), Combining predicate-argument structure and operator projection: Clause structure in Role and Reference Grammar, in *Proceedings of the 13th International Workshop on Tree Adjoining Grammars and Related Formalisms*, pp. 61–70, Association for Computational Linguistics, Umeå, Sweden.

Laura KALLMEYER, Rainer OSSWALD, and Robert D. VAN VALIN, JR. (2013), Tree wrapping for Role and Reference Grammar, in Glyn MORRILL and Mark-Jan NEDERHOF, editors, *Formal Grammar 2012/2013*, volume 8036 of *LNCS*, pp. 175–190, Springer.

Makoto KANAZAWA (2016), Multidimensional trees and a Chomsky-Schützenberger-Weir representation theorem for simple context-free tree grammars, *Journal of Logic and Computation*, 26(5):1469–1516.

Tibor KISS (2005), Semantic constraints on relative clause extraposition, *Natural Language & Linguistic Theory*, 23(2):281–334, ISSN 1573-0859, doi:10.1007/s11049-003-1838-7, `https://doi.org/10.1007/s11049-003-1838-7`.

Anthony S. KROCH and Aravind K. JOSHI (1987), Analyzing extraposition in a Tree Adjoining Grammar, in Geoffrey J. HUCK and Almerido E. OJEDA, editors, *Syntax and semantics: Discontinuous constituency*, pp. 107–149, Academic Press, Inc.

Christian LEHMANN (1982), Directions for interlinear morphemic translations, *Folia Linguistica*, 16:199–224.

Mitchell MARCUS, Grace KIM, Mary Ann MARCINKIEWICZ, Robert MACINTYRE, Ann BIES, Mark FERGUSON, Karen KATZ, and Britta SCHASBERGER (1994), The Penn Treebank: Annotating predicate argument structure, in *Human Language Technology: Proceedings of a Workshop held at Plainsboro, New Jersey, March 8-11, 1994*, pp. 114–119, Association for Computational Linguistics, `https://www.aclweb.org/anthology/H94-1020`.

James D. MCCAWLEY (1998), *The syntactic phenomena of English*, University of Chicago Press, 2d edition.

Stefan MÜLLER (2004), Complex NPs, subjacency, and extraposition, *Snippets*, 8:10–11, `https://hpsg.hu-berlin.de/~stefan/Pub/subjacency.html`.

George ORWELL (2000), *1984*, Ullstein, German translation by Kurt Wagenseil.

Rainer OSSWALD and Laura KALLMEYER (2018), Towards a formalization of Role and Reference Grammar, in Rolf KAILUWEIT, Eva STAUDINGER, and Lisann KÜNKEL, editors, *Applying and expanding Role and Reference Grammar*, pp. 355–378, Albert-Ludwigs-Universität, Universitätsbibliothek, Freiburg.

Owen RAMBOW (1994), *Formal and computational aspects of natural language syntax*, Ph.D. thesis, University of Pennsylvania.

John Robert ROSS (1967), *Constraints on variables in syntax*, Ph.D. thesis, Massachusetts Institute of Technology MIT.

Wolfgang STERNEFELD (2008), *Syntax. Eine morphologisch motivierte generative Beschreibung des Deutschen, volume 1*, Stauffenburg, Tübingen, 3rd revised edition.

Jan STRUNK and Neal SNIDER (2008), Extraposition without subjacency, slides, DGfS 2008, Bamberg. `https://www.hlp.rochester.edu/~nsnider/pubs/ExtrapositionWithoutSubjacencyDGfS2008.pdf`.

Jan STRUNK and Neal SNIDER (2013), Subclausal locality constraints on relative clause extraposition, in Gert WEBELHUTH, Manfred SAILER, and Heike WALKER, editors, *Rightward movement in a comparative perspective*, volume 200 of *Linguistics Today*, pp. 99–143, John Benjamins, Amsterdam.

Robert D. VAN VALIN, Jr. (2005), *Exploring the syntax-semantics interface*, Cambridge University Press.

Robert D. VAN VALIN, Jr. and Randy LAPOLLA (1997), *Syntax: Structure, meaning and function*, Cambridge University Press.

K. VIJAY-SHANKER and Aravind K. JOSHI (1988), Feature Structures Based Tree Adjoining Grammar, in *Proceedings of COLING*, pp. 714–719, Budapest.

Heike WALKER (2017), *The syntax and semantics of relative clause attachment*, Ph.D. thesis, Johann Wolfgang Goethe-Universität zu Frankfurt am Main.

XTAG RESEARCH GROUP (2001), A Lexicalized Tree Adjoining Grammar for English, Technical report, Institute for Research in Cognitive Science, Philadelphia.

Laura Kallmeyer
0000-0001-9691-5990
kallmeyer@phil.hhu.de

Institut für Sprache und Information,
Heinrich-Heine Universität Düsseldorf

Laura Kallmeyer (2021), *Extraposed relative clauses in Role and Reference Grammar. An analysis using Tree Wrapping Grammars*, Journal of Language Modelling, 9(2):225–290
https://dx.doi.org/10.15398/jlm.v9i2.255

This work is licensed under the *Creative Commons Attribution 4.0 Public License*.
http://creativecommons.org/licenses/by/4.0/

Against strict headedness in syntax

Timm Lichte
University of Tübingen

ABSTRACT

Strict headedness is a common idealization in the structural analysis of linguistic entities, particularly in syntax. This contribution takes a critical look at its premises and applications by demonstrating the surprising sloppiness of both the defining concepts and the test procedures, and by showing how strict headedness is nevertheless implemented as an important axiom into virtually all mainstream grammar formalisms. Subsequently, I present a non-trivial head-agnostic analysis based on Tree Unification & Constraints (TUCO) in order to show that there actually is a choice and that strict headedness can be avoided in principle.

Keywords:
grammar, syntax, head, government, dependency, valency, tree unification, tree constraints

INTRODUCTION 1

Headedness can be seen as a partial relation between several entities (phonemes, morphemes, words, phrases, etc.) in a complex linguistic structure that yields a distinction between heads and non-heads. In a more constrained reading, furthermore, a headedness relation is asymmetric (an entity cannot concurrently be the head and the non-head of another entity), single-headed (every entity has at most one head),[1] complete (every entity has at least one head, except for "lexical" entities), and usually also endocentric (the head of an entity is

[1] Headedness with multiple heads will be treated in Section 3.3.

also a component of the entity). I will call this strict headedness (see Section 2), which seems to be the standard conception of headedness.

At least since stucturalist classics such as Jespersen (1924, p. 96), Bloomfield (1935, p. 195), Harris (1951, §16.5), Tesnière (1959), and Lyons (1968, p. 233), (strict) headedness certainly belongs to the set of core notions in mainstream linguistics and is widely used even across otherwise irreconcilable camps.[2] Furthermore, it seems that headedness has developed into a primitive, underived category in language description and modeling, on a par with morphosyntactic categories like case and part of speech, or more syntactic notions such as linearization and linking patterns. Consequently, the identification of "the head" is unanimously seen as very helpful, or downright indispensable, for language analysis, even at a rather descriptive level. For example, the position of a "head" helps typologists to classify languages as "head-initial" or "head-final" (cf. Hoeksema 1992), and it motivates abstract representations assumed by formal syntacticians, be they dependency-oriented or constituency-oriented. It is therefore not surprising that dealing with heads (in particular to identify them) is an important common cornerstone of introductory textbooks and courses (e.g. Kroeger 2005; Radford 2009).

Given the strong, pervasive belief in the necessity of heads, it should be possible to cleanly operationalize what counts as a head and what does not. Interestingly, this is not always as evident as one wishes – at least with respect to syntactic applications. In this contribution, I will point out the inaccuracy of both the defining concepts and the test procedures, which became particularly evident in the context of the Det-or-N debate (e.g. Zwicky 1985; Hudson 1987, 1993; Zwicky 1993; Van Langendonck 1994; Croft 1996; Beavers 2003; Hudson 2004; Müller 2020b). Despite the criticism that strict headedness has rightfully received over the years, it is a surprising and indeed puzzling fact that the notion is nevertheless widespread in syntactic

[2] Sometimes other names are used in place of "head", for example, "primary", "governor" or "functor". Tesnière, at one point, calls it a "mot principal" translating German "Hauptwort" ('main word'), which he says is sometimes used by German grammarians (Tesnière 1959, p. 103). (In fact, German *Haupt* also has the meaning 'head'.) So this is to be understood modulo superficial terminological differences or specific implementations.

theories. I will trace this resistance back to certain basic properties of the formal machinery that is often used to model syntax. Then, I will present an example for a head-agnostic syntactic model using Tree Unification & Constraints (TUCO) as its framework. With the example of long-distance dependencies, it will be shown that TUCO offers enough flexibility and expressive power to immediately capture a wide range of regularities found in syntactic trees without the detour via heads.

THE NOTION OF STRICT HEADEDNESS 2

In this section, I propose an explication of the notions of head and strict headedness, and then discuss the test procedures that are commonly used to distinguish between heads and non-heads. This will be largely based on an influential article by Arnold Zwicky from the 1980s (Zwicky 1985) and the subsequent replies that it provoked. I am not aware of any more recent overview over the topic that is equally detailed and comprehensive.

Notational preliminaries 2.1

In order to make the explication more uniform and crisp, headedness is thought of as a partial, irreflexive relation $<_H$: $\mathscr{P}(E)^+ \times \mathscr{P}(E)^+$ on linguistic entities $E = \{e_1, e_2, ...\}$, where $\mathscr{P}(E)^+$ is the power set of E without the empty set.[3] Since we are concerned with syntax, I assume that E denotes the set of word tokens in a sentence. Then $\mathscr{P}(E)^+$ is the set of possible constructs, that is, a CONSTRUCT is a set of linguistic entities. Constructs will be written using lower-case letters, for example, c, and the set of constructs that is assumed for a sentence is correspondingly denoted by upper-case letter C. The COMPONENTS of a construct c are the largest constructs in C that are different from c,

[3] Note that there is nothing to be said against defining $<_H$ as a total, reflexive relation. Some of the following definitions, in particular the endocentricity and bijectivity idealizations, would have to be adjusted accordingly.

and c is their union. For reasons of simplicity, I will be assuming that the components of a construct are non-overlapping. Whenever a construct has no components, it is called a LEXICAL CONSTRUCT.

To give an example, the sentence *The student ate an apple* could be analyzed as a set of constructs $C = \{\{$*the*, *student*, *ate*, *an*, *apple*$\}$, $\{$*the*, *student*$\}$, $\{$*ate*$\}$, $\{$*an*, *apple*$\}$, $\{$*the*$\}$, $\{$*student*$\}$, $\{$*an*$\}$, $\{$*apple*$\}\}$. Here, $\{$*the*, *student*, *ate*, *an*, *apple*$\}$ would have components $\{$*the*, *student*$\}$, $\{$*ate*$\}$ and $\{$*an*, *apple*$\}$, of which only $\{$*ate*$\}$ would also be a lexical construct.

Following this notation, $c_h <_H c$ means that the construct c_h is a head of the construct c. The transitive closure of $<_H$, $<_H^*$, is called the HEAD PROJECTION. According to standard assumptions, $<_H$ is taken to be endocentric, that is, the constructs that c_h is a head of contain c_h as a component. Idealizations like endocentricity will be addressed in Section 2.4. Following this notational convention, the properties of a construct c are written as P_c.

As it is useful to express the connection between heads and non-heads more directly, avoiding the step "upward" to the level of embedding constructs, I will say that c_h GOVERNS c_i in the construct c, or conversely, that c_i is a DEPENDENT of c_h in c, iff c_i is a component of the construct c and c_h is the head of c (i.e. $c_h <_H c$), and $c_i \neq c_h$. There is thus a government relation $<_G$, which is uniquely defined in terms of $<_H$: $c_h <_G c_i$ iff $c_h <_H c$, and c_i is a component of c and $c_i \neq c_h$. In other words, $<_G$ shares the domain with $<_H$ but not the range.

Going back to our example above with the sentence *The student ate an apple*, one could assume a head relation $<_H = \{(\{$*ate*$\}, \{$*the*, *student*, *ate*, *an*, *apple*$\})$, $(\{$*student*$\}, \{$*the*, *student*$\})$, $(\{$*apple*$\}, \{$*an*, *apple*$\})\}$ such that $\{$*ate*$\}$ governs $\{$*the*, *student*$\}$ and $\{$*an*, *apple*$\}$, etc.

2.2 *Popular definitions*

In his seminal article on heads, Zwicky (1985) illustrates the common intuition about heads in the following way:

> The intuition to be captured with the notion head is that in certain syntactic constructs one constituent in some sense 'characterizes' or 'dominates' the whole. (Zwicky 1985, p. 2)

As with every intuition, however, there are "many directions" into which this intuition can evolve once one tries to make it more explicit. Zwicky himself provides five possible definitions of head-related "dominance"; Hudson (1987, Table 4) later lists eight. Let's go through some of them very briefly.

Morphosyntactic locus 2.2.1

Zwicky's favored definition is based on the distribution of "morphosyntactic marks", that is, potentially visible inflectional properties, which determine what Zwicky calls the MORPHOSYNTACTIC LOCUS (Zwicky 1985, §2.1.3). Using the notation above, we can reformulate Zwicky's definition in the following way:

DEFINITION 1 **Head as morphosyntactic locus** *Given a construct* c_h *and a construct* c *with inflectional properties* $P^{infl}_{c_h}$ *and* P^{infl}_{c}*, then* $c_h <_H c$ *iff* $P^{infl}_{c_h} = P^{infl}_{c}$.

Thus, P^{infl}_{c} is the set of inflectional properties of the construct c that *might* influence the syntactic relations that c can have to other constructs. Zwicky considers two such syntactic relations: agreement and argumenthood. For agreement, he gives the example of the construct *the child*. From Definition 1 it follows that its head is *child,* because it contributes the marking for singular that might participate in an agreement relation with a verb. Conversely, the head of *is controlling those penguins* should be the auxiliary *is*, and the head of *controls those penguins* should be the verb *controls*, since both carry singular marking, which is critical for establishing an agreement relation with the subject.

Turning to the syntactic relation of argumenthood, Zwicky argues that the morphosyntactic locus of a prepositional phrase like *of the news* is the preposition *of* rather than the NP, for the preposition sometimes marks "particular syntactic arguments of the verb" such as in *inform Sandy of the news*. Thus, according to Zwicky, it follows by "analogy", that the head of *every* instance of an P + NP construct is the preposition, even if it is not participating in a syntactic argument relation in the sentence.

2.2.2 Semantic locus

Another head definition that Zwicky discusses is based on the semantic interpretation of a construct (Zwicky 1985, §2.1.1). Here, the construct is supposed to describe "a kind of the thing" that the head describes.[4] One could accordingly call the head the SEMANTIC LOCUS of the construct, and define this very similarly to the definition of morphosyntactic locus:

DEFINITION 2 **Head as semantic locus** *Given a construct* c_h *and a construct* c *with ontological properties* $P^{ont}_{c_h}$ *and* P^{ont}_{c}*, then* $c_h <_H c$ *iff* $P^{ont}_{c_h} = P^{ont}_{c}$.

Thus, P^{ont}_{c} and $P^{ont}_{c_h}$ are supposed to only include semantic properties that somehow pertain to the ontological type of the construct meaning, that is, what it actually denotes. For example, according to Zwicky, the head of the construct *those penguins* is *penguins*, because it "describes a kind of penguin". Similarly, the head of *will leave*, namely *leave*, contributes the "kind" of the event that is described by the construct. While this is all rather vague, Zwicky tries to reify the relevant semantic properties by means of the functor-argument distinction. But this, of course, only shifts the problem to the question what a functor and an argument are supposed to be, even though Zwicky seems to assume that this is independently assured knowledge.

2.2.3 Subcategorization locus

The third head definition on my list is concerned with the morphosyntactic constraints that a lexical item can impose on its context. Zwicky here employs the notion of subcategorization, defining the head as the "subcategorisand". For instance, the verb *give* must "occur with either NP NP or NP *to* + NP as its sisters", whereas the verb *donate* only cooccurs with NP *to* + NP. Both *give* and *donate* are therefore supposed to

[4] See also the extensive discussion of a semantic definition of heads in Croft (1996, §4, §6). Croft proposes a refinement under the notion of a "Primary Information-bearing Unit (PIBU)", which he claims to be superior to morphosyntactic locus.

be the head of the respective, instantiating constructs.[5] What sets this head definition apart from the first two is that it is introversive: the subcategorization properties do not get projected to the headed construct in the sense that the construct then has the subcategorization properties of the head; instead, the subcategorization properties of the head are a subset of the syntactic properties of the headed construct. Let us frame this using our notational idiom:

DEFINITION 3 **Head as subcategorization locus** *Given a construct* c_h *with subcategorization properties* $P^{subcat}_{c_h}$ *and a construct* c_i *with syntactic properties* $P^{syn}_{c_i}$, *then* $c_h <_H c$ *iff* $P^{subcat}_{c_h} \subseteq P^{syn}_{c_i}$ *and* $c_i \in c$ *and* $c_h \neq c_i$.

The problem with this definition, as well as with the underlying notion, is that it is systematically non-functional in the sense that both the verb and the NP sisters can be treated as the subcategorization locus. Zwicky circumvents this to some extent by restricting the subcategorization locus to lexical categories, hence, to the verb in the above cases. But it does not always converge like this. If a construct consists of just two lexical entities, such as determiner-noun constructs, the issue resurfaces. This can be seen from one example in Zwicky (1985, pp. 5–6), namely the construct *each penguin*, where both *each* and *penguin* could be seen to subcategorize for the other: *each* requires a singular count noun, and *penguin* requires *each* rather than *many* or *much* as a determiner. Zwicky avoids this indeterminacy by stipulating, on admittedly "theory-specific" grounds, that *penguin* in *each penguin* is in fact non-lexical in the sense that it is embedded in a phrasal category "Nom". This is a symptom of the general disadvantage of understanding heads as subcategorization loci: due to being introversive, this head notion depends on specifically delimited, nested constructs in order to scale. One telling example is the following *that*-clause, which Zwicky mentions as an instance of a "Comp + S" construct:

(1) that the penguins are flying

Following Zwicky, the head of (1) is the complementizer *that*, because it is lexical and requiring a finite clausal sister. However, the remain-

[5] This view seems to coincide with what is called "Subklassenspezifik" in the German literature. See, for example, Vater (1978), Jacobs (1994, p. 26), Ágel (2000, p. 187), and the summary in Lichte (2015, §2.2.4).

ing words of the clause do not count as lexical here. Instead, they make up an embedded construct (of type S) with a separate head (be it *are* or *flying* or *are flying*). In fact, if the construct in (1) was flat, each of its components would be a head of this construct because it could be argued (e.g., by coming up with adequate minimal pairs) that all of them restrict their morphosyntactic context in some way. This is certainly unwanted if one adheres to the idealization that there is only one head per construct (see Section 2.4). Hence, identifying a head as subcategorization locus presupposes the existence of a very specific structure of constructs in order to be reasonably applied, namely one where only one of the components is a "lexical category".

2.2.4 Government, concord determination, etc.

Zwicky (1985) distinguishes two more allegedly independent head notions that, in my opinion, do not deserve this status. One is "syntactic government", which can be easily included in the definition of subcategorization locus – something Zwicky rejects without adequate justification:

> Syntactic government, speaking rather loosely, is the selection of the morphosyntactic shape of one constituent (the GOVERNED, or SUBORDINATE, constituent) by virtue of its combining with another (the GOVERNOR). Governors are thus easily confused with subcategorisands. Intuitively, the difference is that subcategorization concerns the very possibility of one constituent's combining with some other co-constituent(s), while government concerns the form that a co-constituent has in such a combination. (Zwicky 1985, p. 7)

In other words, Zwicky differentiates between constraints on form (= government) and constraints on existence (=subcategorization) that the head of a construct may impose on non-heads. Yet the examples for subcategorization that Zwicky provides, some of which I mentioned above, are often found together with constraints on form, for example, on the lexical form of a preposition, on the finiteness of the verb, or on the number marking of the noun. So it is not clear why the two notions could not be safely merged.

The other alleged head notion is "determinant of concord", and it looks very similar to government in that it involves constraints on form. But here the scope is narrower, namely on "concord features", that is, features that are subject to agreement, such as number agreement of subjects and finite verbs. The issue with subject-verb agreement – as well as with any other case of agreement, I suppose – is that it is notoriously unclear which is the "determinant": the subject, or the verb. Zwicky (1985, p. 9) argues, by looking at Swahili and aiming at typological uniformity, that in English the subject should be seen as the determinant.[6] Be that as it may, this can be seen as a subcase of government where headedness is particularly hard to decide on if one is seeking for a single head. I will come back to this in the next section when dealing with headedness idealizations.

Of course, the literature holds a plethora of further definitions of a syntactic head, but I claim that their essence is covered by Definitions 1–3. Hudson (1987), for example, adds three definitions to Zwicky's five: head as distributionally equivalent to the construct, head as an obligatory component of the construct, and head as a "ruler". The first two are indeed treated as "operational criteria" in Zwicky (1985, §2.5) that can be, to some extent, related to the head definitions above. They will be addressed in Section 2.3. The term "ruler", on the other hand, is used in dependency theory and largely coincides with what Zwicky calls a head, and is similarly vague (Hudson 1984, p. 78, Zwicky 1985, §2.6). Therefore, it is not really helpful when trying to elucidate what a head is.

Popular test procedures 2.3

As for the three information-based head notions discussed in the previous section, corresponding test procedures straightforwardly suggest themselves: just sort out the source of the important morphosyntactic, semantic and subcategorizational properties of a construct, and this will be the head. However, it does not seem obvious how to operationalize these notationally straightforward tests in a uniform and precise way, that is, how to determine what sort of information came from where.

[6] See Müller (2015, §2) for a general criticism of arguments that rely on typological uniformity.

Maybe for this reason, there are a handful of further, more widely used test procedures for headedness that instead rely on grammaticality (or acceptability) judgments. Zwicky (1985, §2.5) confines himself to two of them, admitting that "they appear to be imperfect guides to the heads in syntactic percolation": (i) a test for "distributional equivalence" and (ii) a test for obligatoriness. This is remarkable because it means that there cannot be a one-to-one relationship with the five head definitions that Zwicky mentions (which is also trivially true, but to a lesser extent, for the selection of three distilled above). And this is critical, because it means that there is actually no way to fully test head notions against each other. Moreover, I will argue that even establishing some one-to-one relationship is difficult, which casts much doubt on this entire approach to head identification.

2.3.1 Substitution test for distributional equivalence

The idea that the head ought to be distributionally equivalent to the governed construct goes back to the structuralist literature (see Zwicky 1985, p. 11 for some references). The rationale is "that the head characterizes a construct in the sense that it is the one constituent that belongs to a category with roughly the same distribution as the construct as a whole" (Zwicky 1985, p. 11). Therefore, replacing the construct with the head should retain the grammaticality (or acceptability) of the sentence and the morphosyntactic "category" of the construct while at most reducing the set of its semantic properties. Using our notational idiom, this can be written down in the following way:

TEST 1 Substitution of the construct *Given a grammatical sentence S comprising a construct c with a component* c_h*,* c_h *is the head of c if* c_h *can be substituted for c in S such that the resulting sentence remains grammatical and compatible with S in terms of morphosyntactic and semantic properties.*

That is to say, *students* in *the students are waiting for the hungry teacher* passes the substitution test for being the head of *the students* since *students are waiting for the hungry teacher* is grammatical and formally and semantically compatible to the extent that the resulting semantics entails the original one. Contrary to this, *hungry* cannot replace *for the hungry teacher* in a similar way. Despite the grammaticality of the resulting sentence, namely *the students are waiting hungry*, the morphosyntactic and semantic properties clearly diverge.

Looking at Zwicky's above-cited rationale, one is inclined to think that the substitution test relies on the morphosyntactic properties of the head, and that it is therefore a test for the morphosyntactic locus. However, this is somewhat speculative as the morphosyntactic properties are actually never spelled out during the test. The same is true of semantic or subcategorization properties. So it is simply not clear what exactly determines substitutability. Therefore, Zwicky even goes so far as to claim that "the distributional equivalent represents a genuinely new head-like notion" (Zwicky 1985, 13).

Another problem, particularly when subscribing to strict headedness, is – and actually this has long been acknowledged[7] – that the substitution test may work in specific cases, but not in general. One obvious problem is the P-NP construct in English, which generally cannot be replaced by either P or NP. Similarly, Det-N constructs only pass the substitution test if N is not a singular count noun. The substitution test may be inconclusive in that both the determiner and the noun pass it, which may happen with demonstrative determiners, for example. Note that this is easy to reproduce, which raises the question of why linguists would still want to rely on it. The answer is: they do not, at least at the token level. Linguists like Zwicky focus on abstract phrase structure rather than token-instantiated strings, and they apply some sort of preselection based on vague statistical and/or theoretical grounds. For example, Zwicky argues that, since Det-N constructs and their N component have "roughly the same" distribution, the N component should be seen as the head. For the same reason, namely the distribution being "roughly the same", the head of Aux-VP constructs is claimed to be Aux. Similarly, S is given head status in Comp-S constructs.

This sort of cherry-picking weakens the importance of the substitution test considerably. Eventually, one remains free to treat it as one of several pieces of evidence in favor, or against, a certain head/nonhead partition. Unfortunately, a similar methodological flaw can be observed in the use of the equally popular omission test, to which we now turn.

[7] Zwicky cites a critical discussion in Lyons (1977) regarding the head of Det-N constructs.

2.3.2 Omission test for obligatoriness

The omission test mirrors the substitution test in that the non-heads are now substituted for the construct, albeit with a negative expectation. Zwicky himself notices that "this criterion is closely related to the preceding one, and might be considered to be an extension of it to (some) syntactically exocentric constructions" (Zwicky 1985, p. 13). In fact, the definition of the omission test looks almost identical to the definition of the substitution test:

TEST 2 **Omission of the head** *Given a grammatical sentence S comprising a construct c with a component c_h, c_h is the head of c if c_h cannot be omitted in S such that the resulting sentence remains grammatical and compatible with S in terms of morphosyntactic and semantic properties.*

Again, it is necessary to add the compatibility condition in order to avoid comparing apples and oranges. For example, the omission of *for* in *the students are cooking for the teacher* would be grammatical, but it would also effect a considerable change in the semantics. Note that the omission is performed piecewise, that is, omitting the whole construct *c* is not possible as the head relation is irreflexive (see Section 2.1).

Even though the expectation might be that the omission test is supplementing the substitution test, their results diverge greatly in many cases: while there are no heads in P-NP constructs following the substitution test, the omission test identifies two heads; conversely, when the substitution test identifies two or more heads, there can be no head following the omission test. Only if there is exactly one head following both the substitution test and the omission test, do the two tests converge (see also Section 2.5.1).

Interestingly, Zwicky (1985, p. 14) arrives at a completely different conclusion, namely that both tests are tests "for the same notion", and that they thus greatly converge. How could that happen? As with the substitution test, one trick is to impose some additional, more theory-driven conditions. For example, to rule out the omission of V in V-NP constructs, which is possible in gapping constructions such as *I ate sushi, and Kiyoko a hamburger*, Zwicky requires that omission be restricted to cases of "optionally present" components, excluding "elliptical" ones. This distinction, however, is not at all trivial both theoretically and methodologically, and moreover touches upon a whole new aspect, namely interpretation in context. Dubious as it might be, this distinction helps Zwicky to identify those parts of a construct as heads that are also selected by the (modified) substitution test. To give another example, Zwicky claims that the omission test supports the view that N is the head in Det-N constructs, because he considers the omission of N elliptical, in contrast to the omission of Det, which is supposed to be optional. The second argumentative strategy that Zwicky applies, and which we have already seen above with the substitution test, is to find positive evidence for a token, postulate it for its type, and then to postulate it for other tokens of the type – even if they do not pass the test. A case in point is the head analysis of NP-VP constructs (Zwicky 1985, p. 13). It is argued that the head is VP because (i) omitting NP is ellipsis, and (ii) VP can be standalone in some cases, when forming an imperative sentence. With this sort of argumentation, we arrive at the curious situation that it "follows" from the omission test that the head of *I ate sushi* is *ate sushi*, even though it does not pass the omission test.

Another more conceptual issue of the omission test is that it conflates two notions of obligatoriness that correspond to either heads or non-heads. Obligatoriness can be attributed to the central role of the head in contributing morphosyntactic or semantic properties. But obligatoriness may also hint at a non-head, namely when being the obligatory argument of the head by virtue of subcategorization properties.[8] Therefore, to be fruitfully applied, the results of the omission test must be set against the subcategorization properties of one of the putative heads.

Strict headedness and other popular idealizations 2.4

One central issue when using the test procedures mentioned, be they information-based or grammaticality-based, is that their use often comes with certain strong expectations of how the headedness relation behaves structurally. Unquestionably, the most significant expectation can be referred to as STRICT HEADEDNESS, namely: each non-lexical

[8] Therefore, the omission test is also popular in valency theory as one criterion for distinguishing arguments from adjuncts. See, for example, Somers (1984), Storrer (1992, p. 105), Jacobs (1994), and Mel'čuk (2004, p. 266).

construct contains exactly one head. In fact, this expectation looks so natural and is so firmly implemented in the formal machinery of many grammar models (see below in Section 3) that I rather want to call it an IDEALIZATION in the sense of Stokhof and van Lambalgen (2011). That is to say, strict headedness is not just the result of temporarily neglecting some "parameters", which would amount to what Stokhof and van Lambalgen (2011) call abstraction. It is an indispensable limitation as to how the data are perceived and how the theory is designed. It is an axiom. In this section, I will try to give a more precise characterization of strict headedness and other idealizations that target the head relation, while adhering to the notational conventions laid out above in Section 2.1.

One fundamental idealization is ENDOCENTRICITY, which basically says that the head is contained within a construct:

IDEALIZATION 1 **Endocentricity** *A head relation $<_H$ must be* ENDOCENTRIC*, that is: if $c_h <_H c$, then c_h is a component of c.*

On the other hand, with the definitions presented here, EXOCENTRICITY manifests as constructs *without* head, since the head relation is deemed irreflexive, that is, a construct cannot be the head of itself. From this it also follows that, at least from the perspective of syntax, lexical words are exocentric by definition.

While endocentricity is generally considered the normal case in syntactic theory, exocentric analyses are also discussed for certain phenoma. Zwicky mentions, among others, the notorious example of sentential constructs consisting of a subject and a verbal phrase, hence NP-VP constructs.[9] Here, it is sometimes assumed that the sentential category emerges from the construct as a whole rather than from the NP or VP alone, for example in Lexical Functional Grammar, assuming an exocentric category S (cf., Bresnan *et al.* 2016, §6.3). Following Zwicky, this assumption is fed by the observation that NP-VP constructs have a "unique distribution" (Zwicky 1985, p. 12) different from both NP and VP. Also, saying that the nominative case of the

[9] Following Zwicky (1985, fn. 9), P-NP and Comp-S constructs can be considered exocentric to some extent, too. Moreover, exocentric, that is, "non-headed" analyses have been proposed for coordination constructions and relative clauses – see, for example, Abeillé and Chaves (2021) and Müller (1999).

NP is governed by the VP is "counterintuitive" (Zwicky 1985, fn. 5). On the other hand, what seems to speak for an endocentric treatment is that the morphosyntactic locus of NP-VP constructs rather lies in the VP, as it contributes (at least in some languages, which might be taken as an argument due to the intended typological uniformity; see the case of agreement in Section 2.2.4) various features, for example, tense, aspect, mood, etc. (Zwicky 1985, p. 6). This contrast has been famously addressed in Government and Binding theory with the introduction of an abstract head INFL, which ultimately helps to establish an endocentric phrase structure for NP-VP constructs.[10]

Another important idealization, which extends endocentricity, is BIJECTIVITY. It basically states that every construct is the head of at most one construct, and every construct contains at most one head:[11]

IDEALIZATION 2 **Bijectivity** *A head relation* $<_H$ *must be* BIJECTIVE, *that is:*

1. *For every construct* c_h *in the domain of* $<_H$*, there is exactly one construct c such that* $c_h <_H c$.
2. *For every construct c in the range of* $<_H$*, there is exactly one construct* c_h *such that* $c_h <_H c$.

With bijectivity alone, it is still possible to have complex constructs that are lacking a head. This can be prevented by imposing COMPLETENESS on the head relation, in the sense that all constructs that consist of two or more components must have a head.[12] This subset of C (the set of constructs in a sentence) is denoted by $C \setminus L$, where L is the set of lexical constructs with only one component (the lexical construct itself). Note that, usually, lexical constructs are word tokens, but this need not be. Lexical constructs can also be complex constructs comprising more than one word token. Completeness can then be formalized as follows:

[10] See Zwicky (1985, Footnotes 5, 9) for some further remarks on the status and treatment of NP-VP constructs.

[11] Syntactic theories that employ multiple heads in some cases are dicussed in Section 3.3.

[12] The Completeness Idealization corresponds to what Richard Hudson calls the "Non-Dangling Principle", which requires that all words have one governor, or "parent" (Hudson 1994, p. 98, Hudson 1998, (29)).

IDEALIZATION 3 **Completeness** *A head relation $<_H$ on C with lexical constructs L must be* COMPLETE*, that is: For every construct c in $C \setminus L$, there is at least one construct c_h such that $c_h <_H c$.*

With bijectivity and completeness, it is possible to define the idealization of STRICT HEADEDNESS in the following way:

IDEALIZATION 4 **Strict headedness** *A head relation must be bijective and complete, that is:*

1. *For every construct c_h in the domain of $<_H$, there is exactly one construct c such that $c_h <_H c$.*
2. *For every construct c in the range of $<_H$, there is exactly one construct c_h such that $c_h <_H c$.*
3. *The range of $<_H$ is $C \setminus L$.*

In other words, strict headedness amounts to head relations in which every non-lexical construct has exactly one head, and a construct can be the head of at most one other construct. One of the nice consequences of strict headedness is that the inverse of the head relation taken together with the government relation (i.e. the head of a construct "governs" all the other immediate components of the construct) forms a tree-shaped graph on *C*. However, note that if we draw this dominance tree on top of a sentence, there might be crossing branches due to the fact that constructs might be linearly non-contiguous.

In order to avoid crossing branches (at least under endocentricity), a further idealization can be added, namely LINEAR CONTINUITY, which requires that heads and headed constructs are continuous sequences with respect to the linear order of word tokens, that is, substrings of the sentence (which corresponds to *C*). In what follows, I will call the linear order of the word tokens of a construct *c* the LINEARIZATION of *c*:[13]

IDEALIZATION 5 **Linear continuity** *A head relation must be* LINEARLY CONTINUOUS*, that is: for all constructs c_h and c with $c_h <_H c$, it*

[13] At the same time, there is a body of work that fundamentally questions the linear continuity of heads and constructs. See, among others, the discussion in Wells (1947, §5), Curry (1961, pp. 65–66), McCawley (1982), Zwicky (1986), and, again just as an example, the implementation in Kathol (1995).

holds that the linearizations of c_h and c are substrings of the linearization of C.

Finally, it is also popular additionally to require LINEAR ADJACENCY, which states that the head is linearly adjacent to all the non-heads of the construct:[14]

IDEALIZATION 6 **Linear adjacency** *A head relation must be* LINEARLY ADJACENT*, that is: if $c_h <_H c$, then the linearizations of c_h and each of the components of c form a substring in the linearization of C.*

Of course, one can observe further idealization-like restrictions in the literature that are related to headedness. For example, once we assign categorical labels to constructs, we can impose head-related constraints on the distribution of those labels in terms of projectivity[15] and uniformity.[16] This sort of restriction has received a lot of attention in the Generative Grammar literature (see, e.g., Chomsky 2008, 2013). But unfortunately, this is dealt with in a rather technical way that already presupposes a certain head/non-head distinction – one which seems to be established merely by tradition.

Empirical evidence for strict headedness? 2.5

In the preceding sections, we have come across several notions and tests concerning syntactic heads, and one predominant idealization, strict headedness. The question now is: can we empirically verify strict headedness based on some test for some head notion? In many cases, the answer must clearly be no. To show this, one may look at very simple sentences, like the following one from German:[17]

[14] See, for example, the Adjacency Principle in Hudson (1987, p. 127).

[15] Projectivity here means that the category of a construct follows from the category of the head.

[16] For example, a category-driven uniformity restriction on head relations could state that, if c_h governs c_d in some head relation, then something of the category of c_d can never govern something of the category of c_h in any of the head relations.

[17] German is chosen here because it has a much richer morphology than English. Furthermore, the author is a native speaker of German, which is helpful when evaluating the tests.

Table 1: Results of the headedness tests applied to (2) and the constructs ⓪–②

heads of	⓪	①	②
omission test	–	*der*	–
substitution test	*aß*, *der Schüler*, *viele grüne Äpfel*	*der*	*viele*, *grüne*, *Äpfel*
morphosyntactic locus	*aß*	*der*, *Schüler*?	*viele*?, *Äpfel*
semantic locus	*aß*	*Schüler*	*Äpfel*

(2) [[Der Schüler]① aß [viele grüne Äpfel]②.]⓪
the pupil ate many green apples
'The pupil ate many green apples.'

As can be seen from (2), we have to presuppose some constructs, here labeled ⓪, ①, and ②, on which the tests and definitions can operate. The results are shown in Table 1 for the omission and substitution test, and for morphosyntactic locus and semantic locus. If a test identifies no or more than one head, it contravenes strict headedness, which seems to be the case for all but semantic locus. In order to make the results in Table 1 transparent, I will discuss the individual tests and definitions in more detail below. Subcategorization locus is neglected here due to its inherent problems, discussed in Section 2.2.3.

2.5.1 Omission & substitution test

The omission test (Test 2) is based on the assumption that the head cannot be omitted without making either the semantics incompatible or the sentence ungrammatical. In (2), any component of ⓪ and ② can be omitted while leaving the co-components in place. Thus, (2) could be replaced with *aß* ('ate'), *der Schüler* ('the pupil'), *der Schüler viele grüne Äpfel* ('the pupil ~~ate~~ many green apples') etc. in certain contexts so that the modified sentence would retain a compatible semantics. Those contexts could, for example, consist of questions such as *What did the pupil eat?* or *What did the pupil do?* Note that the omission test does not say anything about the context of the sentence, which can therefore be freely chosen. By contrast, the determiner *der* ('the') of ① *der Schüler* does not appear to be omissible in any context.

The substitution test (Test 1) aims at the distributional equivalence of the construct with its head. The results in Table 1 mirror the results of the omission test: ⓪ and ② can be replaced by more than one of its components, while ① can only be replaced by the determiner *der* ('the').

Morphosyntactic locus 2.5.2

The morphosyntactic locus (Definition 1) ideally contains all the inflectional properties of a construct. It is a matter of debate, however, what set of inflectional properties should be taken into account, and it also depends on the language. In what follows, I will only consider the properties for gender (MASC, FEM, NEUT), number (SG, PL), case (NOM, ACC, DAT, GEN), and tense (PRES, PAST).[18]

The first difficulty is to assess the inflectional properties of the entire clausal construct ⓪. Assuming that the number property is SG and the tense property is PAST, because they could be seen to express the finiteness of the clause and therefore to restrict its distribution, the morphosyntactic locus should be assigned to the verb *aß* ('ate'). The subject *der Schüler* ('the pupil'), with which *aß* agrees, however, only bears SG.

With the subject construct ①, the determination of the inflectional properties seems to be easier and I will assume, based on the selection above, that these are MASC, NOM, and SG. However, it becomes more difficult when trying to actually assign these properties to one of the components due to inflectional ambiguity. The determiner *der* ('the') is ambiguous in that it can be also used with properties FEM, DAT|GEN, SG, or properties FEM|MASC|NEUT, GEN, PL. *Schüler* ('pupil'), on the other hand, is ambiguous with respect to number and case, only ruling out GEN + SG and DAT + PL. Thus, the only property that is lexically fixed is the MASC property of *Schüler*, and this could be taken to indicate the morphosyntactic locus of ①. However, once the MASC property is set, the determiner *der* actually specifies two other properties, namely NOM and SG, and now *der* seems to act as the morphosyntactic locus of ①. In other words, it is not obvious how to incorporate the morphological ambiguities of constructs while determining the morphological locus.

The situation in the object construct ② is equally inconclusive. ② has the properties MASC, ACC, and PL. *Äpfel* ('apples') could be taken

[18] Properties for person are omitted here for the sake of brevity as the examples will only include constructs with third person.

to contribute MASC and PL, while *viele* ('many') contributes PL too and restricts the case property to NOM or ACC (*Äpfel* furthermore allows for GEN). The adjective *grüne* ('green'), on the other hand, is ambiguous between SG and PL. Thus, the noun seems to be just slightly more specific in terms of inflectional properties compared to the determiner.

2.5.3 Semantic locus

The only clear support for strict headedness in (2) so far seems to come from semantic locus – but is it always like that? (3) shows that the noun and semantic locus in the object construct ② can be missing, and it is much less clear which of the two remaining components is now the semantic locus:

(3) [[Der Schüler]① aß [viele grüne]②.]⓪
the pupil ate many green
'The pupil ate many green ones.'

Of course, this can be explained away by ellipsis: that is, one could argue that only the "complete" or "reconstructed" sentence should be tested. Hence, the head of the Det-A construct *viele grüne* in (3) could be claimed to be an invisible N, which inherits the semantic weight from its presumable antecedent *Äpfel*. Note however that, depending on the context, N could also be taken to just refer to *Dinge* ('things'), which comes with very few ontological properties and is certainly less informative in (3) than the quantifier *viele* and the color adjective *grüne*.

But even if we put aside N ellipsis for the moment, there are other sentences that seem to challenge the semantic foundation of strict headedness in different ways. In (4), for example, it is not clear whether the preposition *in* is solely responsible for contributing the ontological type as was claimed above: whether it contributes a location reading or a path reading crucially depends on the case of the determiner:

(4) [[Der Schüler]① sprang [in der/die Schule]②.]⓪
the pupil jumped in the.DAT/the.ACC school
'The pupil jumped in/to school.'

Therefore, in cases like this, the semantic locus seems rather to reside in both the preposition and the determiner (or any other component with case marking) than in the preposition alone.

A similar issue arises with multi-word expressions with an idiomatic meaning, where it is difficult to decide which of the components contributes the semantics. An example from German is shown in (5), which involves the multi-word expression *ins Gras beißen* ('die', lit: 'bite in the grass'):

(5) [[Der Schüler]① biss [ins Gras]②.]⓪
the pupil bit in.the.ACC grass
'The pupil died.' (lit. 'The pupil bit in the grass.')

It could be claimed that the general ontological type, namely eventuality, is nevertheless contributed by the verb and not by the PP. But apart from this, dying events and biting events differ considerably (the first one being an accomplishment, the second one an activity). Why does that not count here? In other words, a burning question is where to draw the line between decisive and secondary semantic contributions. There is no easy way to answer this question, as far as I can see.

To avoid such problems, one reviewer suggested determining the semantic locus based on the literal meaning of (5). However, one immediate consequence would be that the idiomatic meaning does not have its own syntactic representation, but adopts that of the literal reading(s). Even though this solution has actually been argued for in psycholinguistic and grammar-theoretical work (cf. Lichte and Kallmeyer 2016), perplexity does not seem a good reason to accept such a severe limitation as a general rule. Moreover, besides the issue of identifying the literal meaning(s), it remains unclear whether this entirely solves the problem as long as the semantic properties that are to determine the semantic locus are not listed.

Interim conclusion 2.5.4

To summarize, none of the test procedures discussed supports strict headedness as is: considerable theory-driven assumptions have to be added in order to make this work out even roughly. Therefore, a last escape hatch is to interpret headedness as multi-factorial notion, that is, to assume that strict headedness arises from a specific combination of those test procedures and the underlying primary notions. In other words, the head could be identified as the construct component that always, or at least most often, occurs in a column in Table 1. Based on this rationale, one could then deduce that the verb is the head of the full sentence, while the head of the subject is the determiner and the head of the object is rather the noun, albeit with a smaller margin. While this actually reflects the state of the infamous Det-or-N debate quite accurately (see Hudson 2004), it is methodologically very questionable, because it is becoming increasingly difficult to reject strict headedness on empirical grounds – and virtually impossible when tests and test results are furthermore non-trivially weighted against each other. In a way, strict headedness then turns into something that is empirically taken for granted.

All this raises the following question: Why do we need strict headedness so badly? My guess is that there are at least three concurrently effective reasons, namely (i) that strict headedness is entrenched by tradition, (ii) that it serves to make syntactic theory more uniform, lean, and hence more elegant, and (iii) that it is enforced by the formal framework. While (i) is certainly the case but without immediate scientific value (though being sociologically important), (ii) is more intricate because it also depends on the choice of the syntactic framework. Therefore, in the rest of the paper, I will concentrate on (iii), and, in the next section, try to show how the formal machinery of syntactic frameworks can point the way toward strict headedness.

3 SYNTACTIC MODELING WITH STRICT HEADEDNESS

When interpreted as algebraic structures, syntactic models have two dimensions that are closely related: (i) the DERIVED STRUCTURE, which is the result of applying operations to elements of the carrier set (i.e. lexical and derived structures), and (ii) the DERIVATION STRUCTURE, which is a record of the operations applied to yield a specific derived structure. Headedness can be reflected in both dimensions, either separately or concurrently, by employing a distinction between heads and non-heads. This distinction can be realized differently, not necessarily leading to strict headedness. In this section, though, I will be concentrating on a representative selection of major syntactic frameworks where this distinction leads to strict headedness. More relaxed but less well-known implementations will be covered in the sections to come. The goal will be not only to show the considerable, paradigmatic commonalities and differences between syntactic models, but also to mark that we actually have a choice. In fact, I claim that the implementation of headedness constitutes another fundamental divide between models of syntax, which is orthogonal to that between generative-enumerative and model-theoretic approaches (Pullum and Scholz 2001), or between lexical versus phrasal approaches (Müller and Wechsler 2014).

Strict headedness in derived structures 3.1

In derived structures, strict headedness appears as the necessity to structurally mark one component of a construct as the head, and the other components as non-heads. For example, given a construct $\{c_1, c_2\}$, either c_1 or c_2 has to act as the head, with the other as non-head. Or, in terms of government, either c_1 governs c_2 or c_2 governs c_1. In other words, the government relation must include all components of a construct and it must be ASYMMETRIC. Note that asymmetric relations in syntactic structures are nothing special per se. One very trivial example for an asymmetric relation is linear precedence, which forms a total order on the word tokens of a sentence. Another one is the subset relation on the set of constructs (C above), which forms a partial order. However, the headedness relation is usually incongruent with the linear precedence relation and, by definition, (properly) embedded in the subset relation.

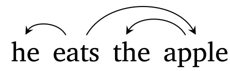
In the remainder of this section, I will give two quite prominent examples from otherwise very distant paradigms that standardly impose strict headedness on their derived structures: dependency structures in Dependency Grammar and phrase structures under the X′-Schema in Generative Grammar.

Dependency structures in Dependency Grammar 3.1.1

Dependency structures are the main theoretical objects in Dependency Grammar (Tesnière 1959; Müller 2020a).[19] While the notion of syn-

[19] Interestingly, the algebraic side of Dependency Grammar, that is, the actual grammar, is usually neglected (but see, for example, Gaifman 1965; Hellwig 2006; Kuhlmann 2010; Müller 2020a, p. 371f).

Figure 1:
Tree-shaped dependency graph



tactic dependency is linguistically congruent with the notion of headedness or government (see van Langendonck 2003), its technical implementation is slightly different: the dependency relation only holds between *lexical* constructs, that is, word tokens. Nevertheless, the dependency relation can be straightforwardly defined based on the headedness relation and the corresponding head projections: given a head relation $<_H$ and word tokens w_1, w_2 and constructs c_1, c_2, w_2 depends on w_1, written as $w_1 \rightarrow w_2$, iff $\{w_1\} <^*_H c_1$ and $\{w_2\} <^*_H c_2$ and $c_1 <_H c_2$. Dependency relations are usually represented as directed graphs where the nodes are word tokens and the edges represent dependencies in such a way that the dependency head dominates (or "points to") the dependent. An example is provided in Figure 1.

As such, dependency graphs do not necessarily implement strict headedness. But they do so with the common set of constraints that are supposed to hold, and which eventually make them dependency *trees* like that shown in Figure 1. Among the most basic, tree-imposing constraints are acyclicity of edges, connectedness of nodes, existence of a unique root, and existence of unique dominance paths (Heringer 1993) – and there is a plethora of further constraints in the literature that build on the more basic ones (e.g. well-nestedness, planarity, projectivity; cf. Maier and Lichte 2011). It should be easy to see that one can deterministically transform a dependency tree into a construct-based head relation which then satisfies strict headedness. Furthermore note that dependency trees are the standard case not only in theoretical work on dependency grammar but also in computational applications that make use of dependencies, for example, in those parsers that follow the guidelines of the Universal Dependencies initiative (de Marneffe *et al.* 2014; Nivre 2015).

That being said, there are actually quite a number of proposals that try to relax the idealization of strict headedness in terms of dependency structures. I will discuss them briefly in Section 3.3.

3.1.2 Phrase structures in Generative Grammar

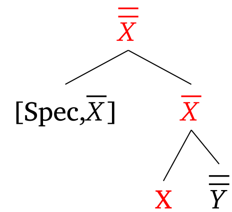
Looking back over time, one might have doubts that phrase structure is adequately characterized as a derived structure. In the early days of Generative Grammar (Chomsky 1957, 1965), when it was called Transformational Grammar and its formal core was considered to be a string-rewriting system, phrase structure had rather more the status of a derivation structure, that is, it served to record which context-free string-rewriting rules had been applied to derive a specific string of words from some start symbol. Now, however, phrase structure is mainly seen as a derivation-agnostic representation of syntactic structure. There are, in fact, grammar formalisms that treat phrase structure as a derived structure in the first place, for example tree-rewriting formalisms such as Tree-Adjoining Grammar (see Section 3.2.2).

As such, phrase structures lack any references to headedness – they just add balanced and labeled brackets to a given string of words. Heads only come into play when restricting the labeling in ways that unequivocally distinguish heads from non-heads. Consequently, only by knowing the labeling rules can one identify heads and non-heads in a given phrase structure.

In this respect, the most influential set of labeling rules is certainly the X′-SCHEMA (Chomsky 1970; Jackendoff 1977; Kornai and Pullum 1990), which goes back to the following famous quote from Chomsky:

> To introduce a more uniform notation, let us use the symbol $\overline{X}$ for a phrase containing *X* as its *head* [emphasis by author]. Then the base rules introducing *N*, *A*, and *V* will be replaced by a schema (48), where in place of … there appears the full range of structures that serve as complements and *X* can be any one of *N*, *A*, or *V*:
>
> (48) $\overline{X} \rightarrow X \ldots$
>
> Continuing with the same notation, the phrases immediately dominating $\overline{N}$, $\overline{A}$ and $\overline{V}$ will be designated $\overline{\overline{N}}$, $\overline{\overline{A}}$, $\overline{\overline{V}}$ respectively. To introduce further terminological uniformity, let us refer to the phrase associated with $\overline{N}$, $\overline{A}$, $\overline{V}$ in the base structure as the “specifier” of these elements. Then the elements $\overline{N}$, $\overline{A}$, $\overline{V}$ might themselves be introduced in the base component by the schema (49):
>
> (49) $\overline{\overline{X}} \rightarrow [\text{Spec}, X]\ \overline{X}$
>
> (Chomsky 1970, p. 210)

Figure 2:
Schema of the X′-Schema
following (Chomsky 1970, p. 210)



This rough sketch of the X′-Schema can be diagrammed as in Figure 2. What is crucial here is that the described labeling restrictions build on the category label of heads (by definition all preterminals in a CFG): the label of a phrase is composed of the label of its head with an added overline or "bar". This headed phrase then projects the category of its head further to the next embedding phrase, namely by adding another "bar" to its category label. Crucially, for each phrase, there is exactly one head or headed phrase that is used in this way. Therefore, even though many important details remain unexplained at least in this sketch (e.g. the exact interpretation of "head" and the treatment of modifiers), the expressed labeling restrictions already bear direct connection to strict headedness, in that every phrase must be labeled in such a way that it participates in exactly one head projection. In other words, every phrase has exactly one head, which must be reflected in the phrase's label. This remains true in later explications and applications of the X′-Schema – see, for example, the "Lexicality" and "Succession" conditions in Kornai and Pullum (1990). Note that Kornai and Pullum (1990) also prove that the use of the X′-Schema has no consequences regarding the set of string languages that can be generated with such constrained CFGs. But that is not what strict headedness is all about. It rather means that it puts an extra burden on grammar writers, who have to decide for each phrase what label to choose in order to keep consistency with the X′-Schema. As we have seen in Section 2.3 and Section 2.5, there is no general and reliable test procedure for this.

3.2 *Strict headedness in derivation structures*

Similarly to derived structures, strict headedness appears in derivation structures as the necessity to treat components of a construct either as head or as non-head. This basically means that each operation that is used for syntactic composition is to be used either with heads or with non-heads, or that the combinatorial operations have dedicated argument positions, which is to say that they are strictly non-commutative. As a consequence, the head/non-head distinction becomes essential for the mechanics of a grammar formalism.[20]

In this section, I will give two examples of derivationally strictly headed grammar formalisms, Categorial Grammar and Tree-Adjoining Grammar, again trying to cover approaches that are as diverse as possible in other respects.

Categorial Grammar 3.2.1

The basic setup of CATEGORIAL GRAMMAR (CG, Ajdukiewicz 1935; Bar-Hillel 1953; Ades and Steedman 1982) is very simple: lexical words are assigned atomic or complex categories such as np/n which may contain slashes that separate input and output subcategories. For example, with category np/n, np is the output and n the input, and the direction of the slash indicates where the input category is to be found. Note that categories are meant to reflect valency properties, if available. Therefore, transitive verbs are usually assigned a category similar to $s\backslash np/np$, which implies that the first input np is to the right, and then another np to the left acts as the second input. Correspondingly, in order to combine these categories, two combinatorial operations are at hand, forward application and backward application, with reference to the direction of the slashes.[21] Schematic examples of forward and backward application are given in Figure 3, using the common proof-theoretic tableau form.

A complete CG derivation then looks like that shown in Figure 4. Proceeding from top to bottom, first the words in the sentence are

[20] Two caveats are in order here. Firstly, headedness in derivation structure does not need to coincide with headedness in derived structure and vice versa. See the example of Tree-Adjoining Grammar below. Secondly, the operations of derivationally strictly headed grammar formalisms could also be used for the representation of other asymmetric relations, such as linear precedence.

[21] Compared to Categorial Grammar, the set of combinatorial operations is considerably extended in Combinatorial Categorial Grammar (Steedman 2000). Nevertheless, the argument made for Categorial Grammar still applies.

Figure 3:
Schematic examples for forward and backward application. σ and τ stand for arbitrary atomic or complex categories

forward application

$$\frac{\sigma/\tau \quad \tau}{\sigma}>$$

backward application

$$\frac{\tau \quad \sigma\backslash\tau}{\sigma}<$$

Figure 4:
Example of CG derivation

he *eats* *the* *apple*

$$\frac{np \quad \dfrac{s\backslash np/np \quad \dfrac{np/n \quad n}{np}>}{s\backslash np}>}{s}<$$

mapped onto their lexically determined categories while retaining the order of the words. Then forward and backward application is used recursively in such a way that only the category *s* remains. As can be seen from this example, the derivation structure is strictly binary (due to the nature of forward and backward application) and the combined categories are consistently treated in a very different way: one being the input of the other. So which is the head? Since categories emerge from valency or subcategorization properties, it seems very natural to identify heads as those categories that "consume" their fellow category in order to satisfy their slashed demand. Thus, *np*/*n* is the head of the construct *np*/*n n*, and every forward and backward application creates another head and non-head pair. In this case, strict headedness is clearly unavoidable.

The strict headedness result for CG can be carried over to similar grammar formalisms such as MINIMALIST GRAMMAR (MG, Stabler 2011), a formalization of Minimalism, even though they differ in some formal details. In MG, lexical words introduce ordered lists of features that can be polarized, which then correspond to the slashed part in CG categories. Instead of forward and backward application, there is only one corresponding operation, namely external merge.[22] However, the arguments of external merge are strictly ordered, the

[22] In Minimalist Grammar and in Minimalism, there is another operation called internal merge or move, which is of a very different nature as it allows for copying and deleting syntactic material.

first one being the head and the second the non-head, which again culminates in strictly headed derivation structures.

Another more surprising connection is proposed by Müller (2013), who stresses the similarity of CG and MG with HEAD-DRIVEN PHRASE STRUCTURE GRAMMAR (HPSG, Pollard and Sag 1994; Müller 2013 [2007]), at least as far as the representation of headedness goes. Of course, HPSG could not be more different in formal terms: it is a constraint-based formalism (with typed feature structures as models) lacking an algebraic structure. In other words, there is no such thing as a derivation in HPSG. Still, Müller (2013, p. 938) claims that "the notation for marking the head of a structure [in MG] […] corresponds directly to the HPSG representation of heads". What he means is that the derivation structures of MG (and also of CG) correspond to the structures of certain syntactic features within the feature architecture of HPSG models. For example, in the HPSG version of Ginzburg and Sag (2001), headed phrases in HPSG carry the features HEAD-DTR and DTRS (with the value of HEAD-DTR being a part of the DTRS list), and there is a head feature principle that ensures the projection of head features from DTRS to the head features of the phrase.

Furthermore, one can often see in HPSG textbooks that almost all phrase types are structured in this way, that is, they are of the type *headed-phrase* (e.g., Müller 2013, p. 195). Thus it might seem that HPSG implements strict headedness in the same way as CG and MG. However, this is wrong in a technical sense. For example, other than CG or MG, HPSG allows for a *non-headed-phrase*, too. Non-headed phrases are commonly used in the analysis of relative clauses (Müller 1999, Müller 2013 [2007], §11.2) or idiomatic constructions (Bargmann 2015).[23] Also, analyses with more than one head are possible – see Section 3.3. Yet, regardless of the technical possibilities, HPSG at its core is "head-driven", that is, designed in such a way that it follows strict headedness as far as possible.

Tree-Adjoining Grammar 3.2.2

Strict headedness is not bound to Categorial Grammar and its more or less direct derivatives; it can also be observed in very unlike but

[23] I am grateful to one of the reviewers for pointing this out.

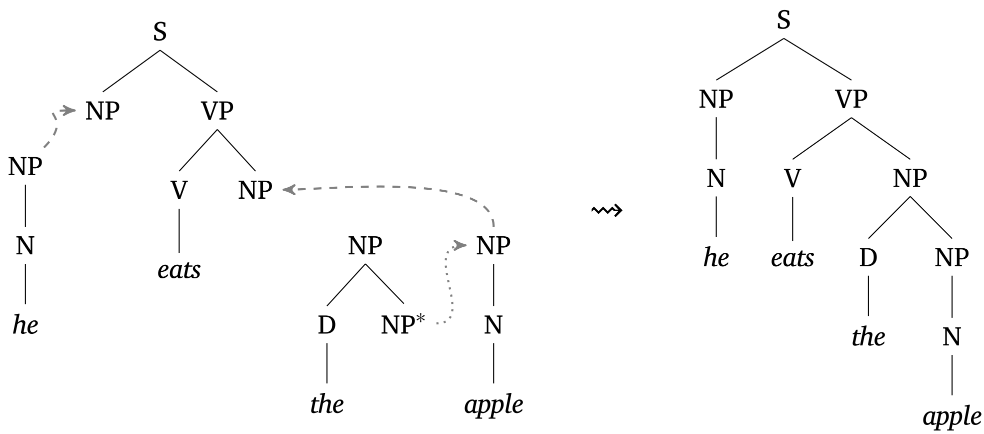


Figure 5: Example of a TAG derivation with the resulting derived tree. The dashed edges indicate substitution, and the dotted edge shows adjunction

still generative-enumerative grammar formalisms too, such as TREE-ADJOINING GRAMMAR (TAG, Joshi and Schabes 1997; Abeillé and Rambow 2000).[24] This is a tree-rewriting formalism where the grammar consists of elementary trees of arbitrary size that can be combined into larger trees using one of two compositional operations, substitution or adjunction: in both cases, a node gets replaced by an elementary tree, but with substitution, it is a leaf node, while with adjunction, it is an internal node. An example is shown in Figure 5, which roughly follows the XTAG Research Group (2001). Thanks to the power of adjunction and the arbitrary size of elementary trees, TAG is said to have an "extended domain of locality" (in contrast to CG and the like) and therefore bears more commonalities with constructionist approaches (Lichte and Kallmeyer 2017).

Despite the constructionist flavor, TAG still adheres to strict headedness, I claim. It is important to note that, for strict headedness, the derived structure on the right side of Figure 5 is not determinative – even though it looks "headed", the derived structure could easily be changed to avoid this impression. What is important, though, is the derivation tree, that is, the nature of substitution and adjunction.

[24] However, despite considerable differences as to formal machinery, TAG is known to have a generative capacity similar to at least some versions of CG and MG (see Joshi *et al.* 1990).

Both operations are non-commutative in the sense that there is a replacing elementary tree and a target elementary tree, and switching those roles leads to different derived trees. Therefore, TAG derivations are usually represented as a derivation tree where the target dominates the substituting or adjoining tree.

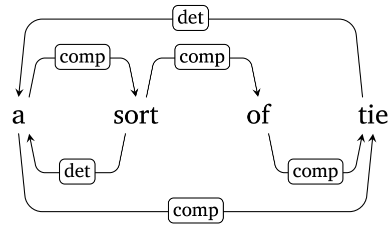
Yet, what is still missing to establish the connection between TAG and strict headedness is a certain interpretation as to what elementary trees represent in syntactic terms. It has long been mainstream to regard elementary trees as realizations of subcategorization properties of the lexical anchor (cf. Abeillé and Rambow 2000; Frank 2002; Lichte 2015, §5.3). That is, the lexical anchor counts as the head of the domain of the anchored elementary tree. Looking back at Figure 5, this is nicely exemplified with the elementary tree of the transitive verb *eats*, which contains NP slots for its subject and object. Moreover, headedness is indicated by choosing a phrase structure that roughly follows the X′-Schema with respect to how the nodes are labeled. It is therefore quite obvious that substitution and adjunction are linguistically interpreted in such a way that they separate heads from non-heads. This culminates in the idea that TAG derivation trees could be homomorphic to canonical dependency trees. Unfortunately, this is not always the case (cf. Rambow *et al.* 1995; Kallmeyer and Kuhlmann 2012).

In a nutshell, with non-commutative operations and the entrenched view of lexical anchors as the heads of the domain of an elementary tree, TAG can be classified as being strictly headed. As a consequence, the grammar writer has to decide for each composition of elementary trees, which to model as the head and which as the non-head – a decision for which, as we have seen in Section 2, there seems to exist no satisfying guidance.

3.3 *Headedness with multiple heads*

The derived and derivation structures that we have seen so far are *mono-headed*, that is, there is at most one head per construct. In fact, mono-headedness is a precondition for being strictly headed because of the Bijectivity Idealization (see p. 305). Yet there are a number of proposals in various frameworks that (more or less explicitly) make use of multiple heads.

Figure 6:
Multi-headed dependency structure
(from Hudson 2004, p. 39)



One prominent variant of Dependency Grammar that explicitly allows for multi-headed syntactic dependencies is Word Grammar (Hudson 1984, 2007). An example from Hudson (2004, p. 39) is given in Figure 6. In this analysis of an intricate Det-N construct, Hudson (2004) proposes that Det and N should be seen as being mutually dependent on each other, avoiding the difficulty of mono-headed dependency analysis.[25] Hence, multiple heads are an interesting tool, particularly in constructs where there is more than one good candidate for the head, or when a construct can be seen to have more than one governor (which are not necessarily components of the same construct). Other possible cases for multiple heads are therefore relative clauses, raising and control constructions, and coordination constructions. In relative clauses, the relative pronoun agrees with the modified noun, but the relative phrase can be seen to be governed from within the relative clause. In raising constructions such as *Kim seems to sleep*, the raised noun *Kim* can be seen to be governed by the raising verb *seems* with regard to agreement and case properties, and by the embedded infinitive verb *to sleep* in terms of semantics. There are also control constructions such as *Kim tries to sleep*, in which *Kim* acts as a semantic argument of both *tries* and *to sleep*. Lastly, multiple governors can be also assumed for the subject in VP coordination constructions such as *Chapman eats cookies and drinks beer* (Sarkar and Joshi 1997).

However, it is important to note that multi-headedness does not coincide with exocentricity, that is, the lack of heads, in which case the head properties of a construct cannot be attributed to any of the

[25] One reviewer pointed out the similarity to the "mutual selection" approach in HPSG (Pollard and Sag 1994, §9.4), in which Det selects N and N selects Det. Despite "mutual selection", however, the Det-N construct is not treated as multi-headed in HPSG, as N clearly acts as the head, and Det as the non-head.

components.[26] Moreover, a certain kind of multi-headedness, which is not intended here, emerges when representing different types of dependencies (e.g. syntactic and semantic) in one dependency structure. This kind of conflation can be observed, for example, in Meaning-Text Theory (Mel'čuk 1988) and Extensible Dependency Grammar (XDG, Debusmann *et al.* 2004).[27] When considering only syntactic dependencies, multi-headedness is banned in such approaches – even though semantic dependencies are treated more liberally in this respect. In fact, the use of multi-headedness in Word Grammar is quite harshly criticized by Mel'čuk (2009, §1.1) as a "confusion between different types and/or levels of dependency" (p. 68).[28]

Of course, multi-headedness has also been discussed as an option outside Dependency Grammar, for example in TAG (cf. Chen-Main 2006) and HPSG (e.g. Abeillé 2003). Furthermore, movement in Generative Grammar can be generally perceived as an operation (or relation) that helps to express multi-headedness (or more precisely multi-dominance) in phrase structure by means of empty categories.[29]

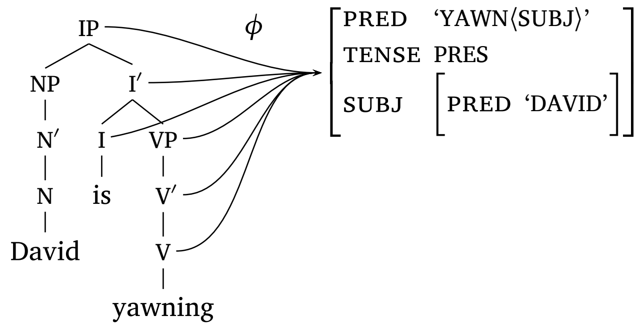
[26] Note that, in the dependency framework, intermediate or partial structures have been discussed that can be considered headless, such as "connection graphs" (Gerdes and Kahane 2013) or Bubble Trees (Kahane 1997). Thanks to one of the reviewers for pointing this out.

[27] See also the discussion of multi-headedness as a result of different sorts of heads in Zwicky (1993).

[28] Mel'čuk's criticism might look surprising given that, in his Meaning-Text Theory, he assumes two syntactic layers, Surface-Syntactic Structures and Deep-Syntactic Structures, which taken together would impose multi-headedness. It seems, though, that Deep Syntactic Structure is rather seen as an interface level between syntax and semantics. This view is also adopted in recent work within computational linguistics in the context of the Sequoia French Treebank (Candito *et al.* 2014; Michalon *et al.* 2016).

[29] Multi-dominance, or "multi-government", arises in configurations where a construct is governed by more than one head. These heads can be (i) components of one construct (in which case there is also multi-headedness), or (ii) components of overlapping constructs so that they govern the common components. What is meant here is multi-dominance in overlapping constructs. A classical example known from GB Theory is the raising construction *John seems t to like ice cream* (Chomsky 1981, (9iv)), in which *John* is said to be governed by both *seems* (in terms of case marking) and *like* (in terms of θ-role assignment). This is captured in GB Theory, following our terminology, by first base-generating *John*

Figure 7:
C- and f-structure of an LFG analysis of the sentence *David is yawning* (copied from (11b) in Müller 2020a, p. 226)



Alternatively, Lexical Functional Grammar (LFG, Kaplan and Bresnan 1982; Bresnan *et al.* 2016) adopts a (more relaxed) version of the X′-Schema, but avoids movement by postulating a flexible mapping ϕ from nodes of the phrase structure (c-structure) to components of a structured functional representation (f-structure). With this ϕ mapping, it is possible to associate several c-structure heads with the same f-structure head, for example, extended projections that span a lexicalized VP and a functional IP. An example of this is shown in Figure 7 (copied from (11b) in Müller 2020a, p. 226). It could be argued (as one of the reviewers did) that I and V act as two functional heads of the sentence *David is yawning*. But, again, this view seems to blend different sorts of heads and representations. When considering c-structure and f-structure separately, each is strictly headed: c-structure largely adheres to the principles of the X′-Schema, and even the exocentric category S is usually governed by an "extended head" (Bresnan *et al.* 2016, p. 136) such as I. The f-structure, on the other hand, can be seen to correspond to dependency structures in which dependency relations hold between the PRED features according to their hierarchical order (Przepiórkowski and Patejuk 2020). Note that PRED basically contains a valency list, and the well-formedness conditions of completeness and coherence ensure that the members of the valency list are realized as siblings of PRED (Bresnan *et al.* 2016, §4.7).

Thus, even if multi-headedness seems a widespread alternative to strict headedness (and similarly exocentricity), it does not fundamentally compromise strict headedness. One reason is that multi-

in the construct of *like* and then moving it to the construct of *seems* indicated by the trace *t*.

headedness nevertheless typically shares many idealizations such as the important Completeness Idealization (see Section 2.4, p. 306), which states that every construct bears a head. Another reason is that multi-headedness is also commonly used as a resort for when strict headedness cannot be easily upheld, but it is not used as a general replacement. However, what is intended in this work is precisely the general elimination of strict headedness.

AN ALTERNATIVE LOOK AT HEADS 4

Before presenting an example of a head-agnostic syntactic model, I would like to briefly clarify the conception of heads that underlies it. This conception does not imply strict headedness and the other idealizations mentioned in Section 2.4. Instead, heads are seen as secondary, following from other more fundamental properties of a construct and its alleged heads. In order to capture this, I propose a CONTRIBUTION-BASED conceptualization: something is a head by virtue of contributing some information (i.e. properties) to the embedding construct. In accordance with the head definitions in Section 2.2, properties can be morphological, syntactic, or semantic. Furthermore, the contribution can be made to the embedding construct as a whole, or to another component of the construct. The following explication is kept as simple and general as possible.

Properties are formally treated as unordered flat PROPERTY NAME SETS (PNS) such as {MASC,NOM,HUMAN}.[30] At this level, no distinction is made between properties from different linguistic domains, so that morphological, syntactic and semantic properties are lumped together. Note, however, that PNS can be easily converted into a more ordered format such as a feature-value structure (also with complex values), and vice versa. In order to make the presentation more readable, descriptions are used rather than fully resolved models. Accordingly, PNS may include Boolean operators such as disjunction (|) and negation (¬) with their usual semantics. On top of that, certain

[30] Property name sets are common, for example, in the Lexicon-Grammar framework (Gross 1994; see also Lichte *et al.* 2019).

natural implications are presupposed, for example that {MASC} implies {MASC,¬FEM,¬NEUT}. However, these notational conventions are by no means essential for the argument made here.

Applying the notation of PNS to the German example (2), here repeated as (6), one could decorate the corresponding construct hierarchy with PNS as in Figure 8. Note that the PNS are chosen to serve the example.

(6) [[Der Schüler]① aß [viele grüne
  the.NOM.SG pupil ate many.ACC green.ACC
  Äpfel]②.]⓪
  apples
  'The pupil ate many green apples.'

Firstly note that the PNS of the two NPs are simply unifications of the PNS of its components. But this does not have to always be the case. For example, the PNS of the sentence is not constructed by unifying the PNS of its components; it only takes over some properties of the verbal component *aß* ('ate'). These shared or CONTRIBUTED PROPERTIES, which are shown underlined in Figure 8, can then be used to determine the contributional head of a construct, namely by taking the number and kind of contributed properties into consideration. For instance, the determiner in the subject NP *der Schüler* ('the pupil') could be argued to be the head due to the number of contributed properties (similar to, e.g., Zwicky 1985 and Hudson 1987), whereas the noun would be the head when counting only the semantic property HUMAN (which is what Croft (1996) would probably argue for). In other words, under a contribution-based determination of heads, the head status of a component depends on the other components of a construct and the measure involved. Consequently, there can be more than one head per construct across and within different measures.

If PNS are used as in Figure 8, one tricky aspect of measuring property contribution is the "horizontal" or "introversive" contribution from one component to the other, but not to the embedding construct. This is usually the case with subcategorization or valency restrictions (see Section 2.2.3) by which, for example, a finite transitive verb like *aß* is taken to impose the obligatoriness and case of subject and object. To make such contributions visible, a position index is added to the property name indicating to which of the construct's

der Schüler aß viele grüne Äpfel
{EVENT,PAST,SG}

- *der Schüler*
  {NOM,MASC,SG,HUMAN,DEF}
  - *der*
    {NOM,MASC,SG,DEF}|
    {DAT|GEN,FEM,SG,DEF}|
    {GEN,PL,DEF}
  - *Schüler*
    {NOM|DAT|ACC,MASC,SG,HUMAN}|
    {NOM|GEN|ACC,MASC,PL,HUMAN}
- *aß*
  {EVENT,PAST,SG,1.NOM,3.ACC}
- *viele grüne Äpfel*
  {ACC,MASC,PL,INDEF,FRUIT}
  - *viele*
    {NOM|ACC,MASC|FEM|NEUT,PL,INDEF}
  - *grüne*
    {NOM|ACC,MASC|FEM|NEUT,PL,INDEF}|
    {NOM|ACC,FEM,SG,DEF|INDEF}|
    {NOM|ACC,MASC|NEUT,SG,DEF}
  - *Äpfel*
    {NOM|ACC|GEN,MASC,PL,FRUIT}

Figure 8: Construct hierarchy with property name sets for *der Schüler aß viele grüne Äpfel* ('The pupil ate many green apples.'). The underlined properties are contributed to the embedding construct

components it is being contributed. For example, 3.ACC means that the third component of the construct must have the ACC property.

5 HEAD-AGNOSTIC SYNTACTIC MODELING

The PNS-enriched construct hierarchy in Figure 8 can be regarded as the essence of the derived structure of a head-agnostic syntax. The question now is how to arrive at such a derived structure while avoiding the pitfall of headed derivations. Remember that this does not exclude non-commutative operations altogether since we still have to deal with asymmetric relations like linear precedence and the embedding of constructs in larger constructs. But, apart from that, the operation to compose the components of a construct and their PNS must be commutative. In what follows, I will illustrate this sort of grammar formalism while trying to keep the example as general and to the point as possible.

5.1 *An example with TUCO*

The simplest approach I can think of is to use trees to express construct embedding, unification to compose these trees, and tree constraints to guide tree unification and to account for linearization patterns. Accordingly, I will call this sort of syntactic framework TREE UNIFICATION & CONSTRAINTS (TUCO). Despite its simplicity and the absence of strict headedness, the claim will be that this framework nevertheless provides sufficient means to formalize natural language syntax.

5.1.1 Elementary structures

To begin with, the TUCO elementary trees representing the lexical entry of the transitive verb *aß* (‘ate’) are shown in Figure 9. In this example, the valency roles are represented as separate trees, one for the nominative NP and one for the accusative NP. Within these trees, nodes are labeled with PNS that contain contributed as well as “governed” properties, for example, the agreement property SG and the

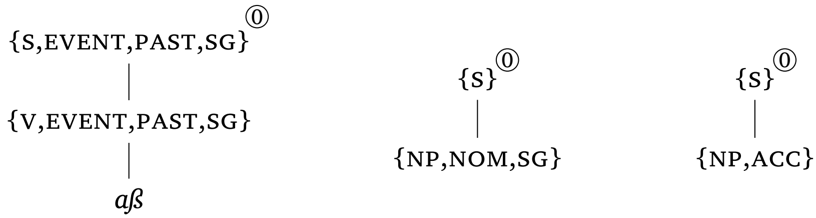


Figure 9: Elementary TUCO trees representing the lexical entry of the transitive verb *aß* ('ate')

case properties NOM and ACC. Moreover, nodes in TUCO trees may carry special marks ⓪, ①, …, which serve two purposes: (i) to constitute a syntax-semantics interface (which is not shown here), and (ii) to express the identity of nodes in the final derived structure. By marking all the root nodes of the trees with ⓪ in Figure 9, it is indicated that these nodes must be combined into one node during the derivation. In other words, even though the valency information of *aß* is distributed over several trees, the mark ⓪ in their respective root nodes ensures that they eventually belong to the identical S node. Note that, as usual, the marks are freshly chosen each time the lexical entry of *aß* is instantiated.

Composition 5.1.2

The composition of TUCO trees uses a sort of TREE UNIFICATION that is akin to the notions in Interaction Grammar (Guillaume and Perrier 2009).[31] A tree is understood as the unique model of a minimal set of tree descriptions, allowing only for descriptions of immediate dominance and immediate precedence relations between nodes. Hence, when unifying trees, one is actually unifying two sets of descriptions, and subsequently compiling all their minimal models – minimal in the sense that no nodes and edges may be added. As for the nodes, tree unification implies the composition of their PNS by set union and the identification of their link marks. To make things easier for now, I will be assuming that nodes with PNS constitute the "non-terminal" nodes, while "terminal" nodes are labeled with word forms written

[31] See also the specific use of tree unification in Popowich (1989), Gerdes (2004), Kahane (2006) and Lichte (2012, 2015). At least the framework presented in the latter work, Synchronous Tree Unification Grammar (STUG), also allows for head-agnostic syntactic modeling.

with italicized font.[32] With this distinction, it is straightforward to impose the usual well-formedness conditions on derived trees, namely that non-terminal nodes must not be leaf nodes, and that terminal nodes are leaf nodes with exactly one non-terminal node immediately dominating them.

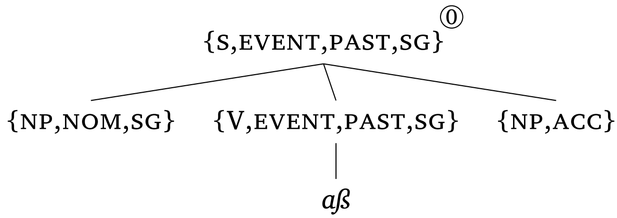
5.1.3 Constraints

When applying tree unification to the lexical trees in Figure 9, there will be several derived trees that are not desirable in linguistic terms. For example, nothing so far prevents unifying all the nodes dominated by the root nodes (which must be unified due to the common marker ⓪), thus resulting in a node with the awkward looking PNS {V, NP, EVENT, PAST, SG, ACC, NOM}. In order to achieve only reasonable solutions, one has to furthermore specify TREE CONSTRAINTS, that is, conditional statements with tree descriptions on the left- and right-hand sides. To give a very simple example, the tree constraint {A} $\Rightarrow$ {A, B} imposes the following: if there is a node with property A, then it also has property B. Conversly, {A, B} $\Rightarrow \perp$ indicates that A and B are incompatible and there is no solution whenever a PNS contains both. Thus stating that {V, NP} $\Rightarrow \perp$ will prevent nodes from carrying the categories V and NP at the same time. Such constraints pertaining to single PNS, as well as more complex tree constraints, are shown in Figure 10, guiding the unification of the lexical trees in Figure 9 in the desired way, that is, unification of nodes is only possible when all tree constraints are fulfilled.

The more complex tree constraints have the same basic shape and semantics as the PNS constraints above, but they describe the configuration of the nodes in a well-formed tree in terms of dominance ($\rightarrow$) and linear precedence ($\prec$). The tree constraint at the bottom of Figure 10 basically states: if there is an S node with three separate daughters with categories NP, V, and NP, then the two NP nodes have to surround the V node. The convention will be that nodes in the description are treated as separate nodes in the model as long as they are not explicitly identified in a description.

[32] One could also generalize PNS to terminal nodes.

Intermediate derived tree:



Tree constraints:

{V, NP} ⇒ ⊥,
{S, NP} ⇒ ⊥,
{S, V} ⇒ ⊥,
{ACC, NOM} ⇒ ⊥,

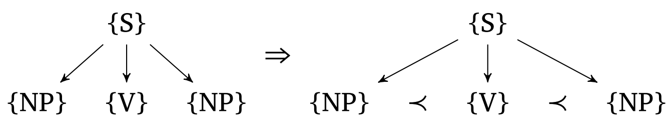


Figure 10:
Intermediate derived tree using the elementary TUCO trees in Figure 9 and conforming to the tree constraints shown below

The full parse 5.1.4

We now have everything to derive the sentence *der Schüler aß viele grüne Äpfel* in the way depicted in Figure 11, while respecting the tree constraints in Figure 10 and Figure 12.[33] The first two tree constraints in Figure 12 guarantee that there is at most one nominative NP and at most one accusative NP under an S node, whereas the last two constraints determine the internal structure of an NP.[34] The resulting well-formed derived tree is shown in Figure 13.

[33] As the constraints in Figure 10 are not specified for case, the alternative OVS word order

(i) Viele grüne Äpfel aß der Schüler.
many.ACC green.ACC apples ate the.NOM.SG pupil
('Many green apples, the pupil ate.')

would also be licensed.

[34] Note that I do not claim this to be a valid generalization about German. Rather, the aim is to demonstrate that it is possible to express this kind of constraint. There are perfectly grammatical sentences in German that contain more than one noun phrase with the same case as immediate components. Thanks to one of the reviewers for pointing this out.

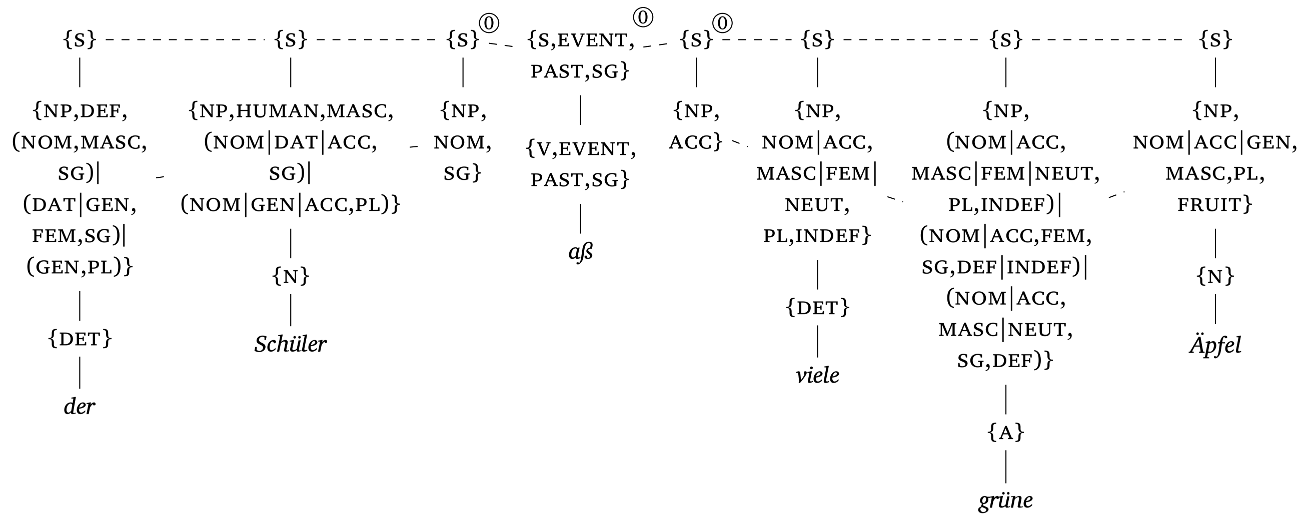


Figure 11: Head-agnostic TUCO derivation of *der Schüler aß viele grüne Äpfel* using, among others, the elementary trees in Figure 9 and complying with the constraints in Figure 10 and Figure 12

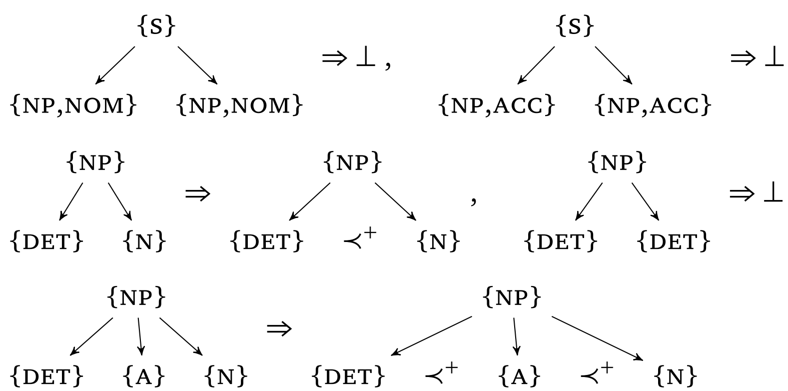


Figure 12: Further TUCO constraints used in the derivation in Figure 11

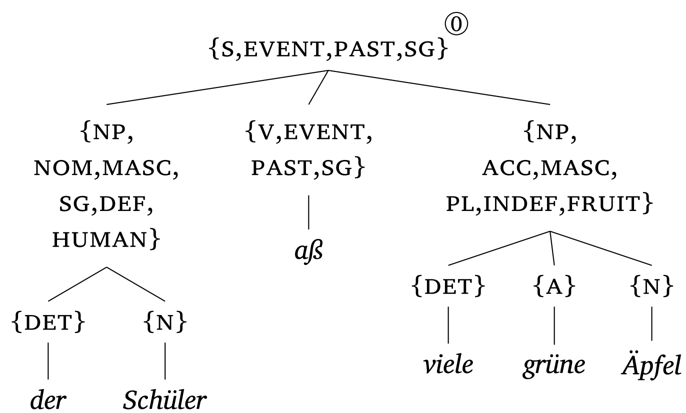


Figure 13: Resulting derived tree of the derivation in Figure 11

Note that both the derived and derivation structure do not presuppose a distinction between heads and non-heads, even though the analysis implements strong lexicalization, including the representation of valency. So it might seem that headedness has slipped in again, since the valency contribution of the verb *aß* contains nodes for the governed nominative and accusative NPs, and it might seem that the derivation hinges on their presence. But this is an illusion. Since valency structures are unified, the same result could be achieved when omitting those nodes in the contribution of *aß*. Moreover, one could *concurrently* assume an entry for subjects that in turn contributes to the preterminal node of *aß*, namely by specifying (or "checking") agreement properties. In other words, there is not necessarily a single head-like component on which the derivation depends; rather, the components of a construct may contribute equally to the syntactic structure of the construct.

5.2 *Notational enhancements*

The example of a head-agnostic analysis presented here uses only the basic elements of the TUCO framework, in particular tree constraints built from an explicit tree description language. While the bare expressive power seems sufficient, the number of tree constraints in a TUCO grammar might very soon become confusingly large. For example, there is one constraint in Figure 12 explicitly prohibiting the occurrence of two DET nodes within an NP, another constraint to determine the order of siblings DET and N, and a third constraint for the correct linearization of DET, A and N. Instead, it would be desirable to have just one concise constraint and description of a well-formed NP.

Fortunately, there are numerous ways in which higher-level abstractions that help to increase the conciseness and readability of a TUCO grammar can be defined from elementary tree constraints. In this section, I will briefly discuss one such abstraction that makes use of bracketing and two-dimensional regular expressions – two-dimensional in the sense that they can refer to both linear precedence and dominance in trees.

Trees can be written as bracketed expressions with the parent immediately after the opening bracket and the children following it. For example, in [{A} {B} [{C} {D}]], {A} is the parent of {B} and {C}, and {C} is the parent of {D}.[35] The linearization and composition of nodes is usually taken literally, that is, the denotation of this expression exactly contains one tree with nodes {A}, {B}, {C}, {D}, and the dominance relations and linear order indicated in the expression.

In the following, however, I will deviate from this convention in two respects: (i) bracketed expressions will be interpreted as constraints by marking the antecedent, that is, the nodes whose existence is presupposed in order for the consequent to apply, with the hash symbol, #; (ii) the string following the parent will be interpreted as a regular expression of children nodes, which fully characterizes the possible strings of children nodes of the given parent. With this adaptation, the structure of NPs in our example can be captured with just one contraint in (7):[36]

(7) [{NP}# {DET}? {A}* {N}]

In prose, the constraint means the following: if a node is an NP, it must immediately and only dominate at most one DET node followed by an arbitrary number (including zero) of A nodes, followed by one N node. To capture this with basic TUCO constraints would require the set in Figure 14. The left constraint states that an NP node may only dominate nodes with properties DET, A or N. The right constraint states that an NP node must dominate one N node at least. The two constraints below furthermore impose that DET must be the first node and N must be the last node under NP. Note that the meaning of the constraint in (7) would change considerably if the antecedent marker # was shifted to DET, for example, or assigned to several nodes at the same time.

[35] The bracket notation of phrase structure trees is widespread in linguistics and goes back at least to Chomsky (1972, p. 130).

[36] This is reminiscent of the use of regular expressions in LFG's c-structure constraints (Kaplan and Bresnan 1982, p. 277), which have the shape of context-free rules. TUCO is more expressive, however, as TUCO constraints may vertically extend beyond the parent and children of a node – see Section 5.2.2 below.

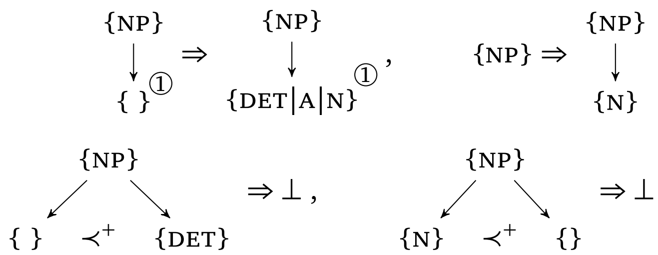


Figure 14: The set of basic TUCO constraints that is equivalent to the single constraint in enhanced bracket notation shown in (7)

5.2.2 Regular expressions of dominance

The Kleene star operator can also be used to describe non-immediate dominance between nodes, similarly to the description of non-immediate linear precedence in (7). The main difference in the notation is that, while in (7) the Kleene star was attached to a child node, it is now attached to a parent as indicated in (8):

(8)
```
[{S,¬SLASH}   [1]{ }
              { }*
              [{S,SLASH}* [{S,SLASH}   { }*
                                       [[1]{ } {TRACE}#]
                                       { }* ]]
              { }* ]
```

The part with [{S,SLASH}* roughly means that there is an arbitrarily long dominance path consisting only of S nodes with a SLASH property. The boxed number [1] is a PNS label that indicates the equality of the two labeled PNS. The equivalent basic TUCO constraint is given in Figure 15. Note that the right-hand side of the constraint spans the entire conjunction including the two implications. This is necessary in order to consistently ensure the presence of properties S and SLASH in the dominance path.

The example obviously hints at filler-trace analyses of long-distance dependencies, as in *what did Kim claim Sandy ate t.* In transformational terms, the constituent *what* is usually said to be base-generated at the position of the trace *t* and later moved to some sentence initial position (see, e.g., Chomsky 1986, Chapter 6). However, TUCO does not have transformations. The SLASH property is therefore named after the slash mechanism in GPSG (Gazdar 1981; Gazdar *et al.* 1985, Chapter 7) and HPSG (Pollard and Sag 1994, Chapter 4), which

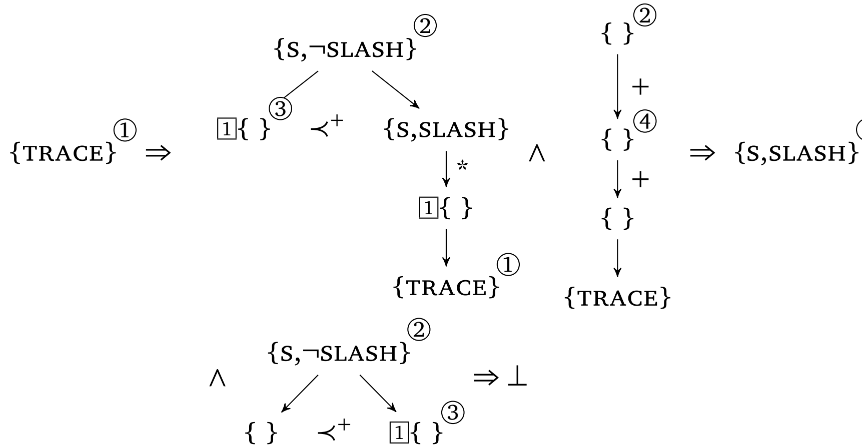


Figure 15: A basic TUCO constraint that is equivalent to the constraint in enhanced bracket notation shown in (8)

copes with these "unbounded dependency constructions" in a non-transformational manner. Similarly, in the constraint above, the first child of {S, ¬SLASH} acts as the filler of the trace under {S, SLASH}.

At least two further constraints that come with filler-trace analyses of this sort are not yet integrated in Figure 15. One is that, under S nodes with a SLASH marking, exactly one child must either contain a trace and carry the SLASH property, or be the trace.[37] Secondly, there has to be exactly one filler for each trace. All this can be captured by adding the single constraint in (9):

(9) [{S,SLASH}# {¬SLASH}*
({S,SLASH} | [{ } {TRACE}])
{¬SLASH}*]

Note that {¬SLASH} cannot dominate a trace without a filler because every TRACE property initiates the projection of the SLASH property up to the filler due to the constraint in (8). Both constraints together license the derived tree for *What did Kim claim Sandy ate* in Figure 16. The underlying elementary trees, from which the tree is assembled

[37] This is probably too strict, but serves the example. Pollard and Sag (1994, Chapter 4) also discuss the possibility of multiple traces in connection with tough movement and parasitic gaps. Gazdar (1981, §3) shows that his slash mechanism is also able to capture Across-the-Board extractions with multiple traces.

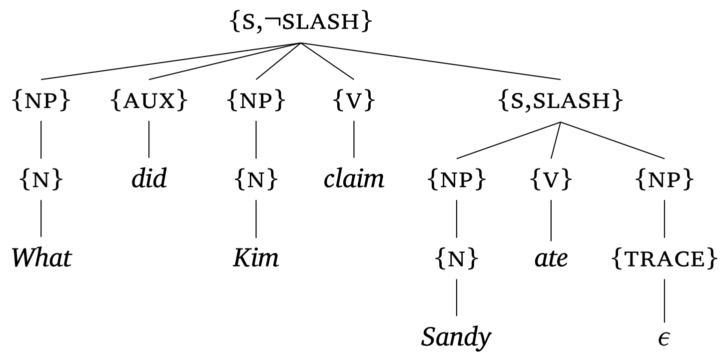


Figure 16: Tree for *What did Kim claim Sandy ate* satisfying the two constraints in (8) and (9)

via tree unification, are not shown here. There are plenty of ways in which this could be done, so I will leave that to the readers.

What should be obvious, though, is that at no point do we need strict headedness in order to arrive at concise theories or intuitive representations. As shown in this section, higher-level abstractions such as two-dimensional regular expressions can be defined based on regular TUCO constraints. They offer enough flexibility and expressive power to immediately capture a wide range of regularities found in syntactic trees without any detour via heads.

6 CONCLUSIONS

In this article, I have tried to delineate one of the most dominant idealizations found in mainstream syntactic theory: strict headedness. The idea that each construct should be partitioned into heads and non-heads is unanimously taken for granted, or so it seems. At the same time, however, the considerable empirical issues as to the operationalizability of the competing head notions are well-known. I hypothesized that this puzzling paradox can be at least partly explained with the formal mechanics of the predominant syntactic models, which presuppose strictly headed derived or derivation structures. Subsequently, I presented a non-trivial head-agnostic grammar formalism based on TUCO in order to show that there actually is a choice that we should be aware of.

Another explanation for the resilience of strict headedness could be the term *head* itself, which gives rise to certain anthropomorphic expectations, as beautifully proven by the following quote from Croft (1996, p. 57):

> For some people, myself included, the two-headed model for phrases and clauses implied by the profile-determinant semantic definition of head is an unsatisfactory conclusion. The idea of a two-headed phrase sounds about as natural as a two-headed baby.

Since no linguist would be cruel enough to deliberately create a "two-headed baby",[38] the goal of producing a one-headed baby is immediately understandable from a psychological point of view. So, maybe, we should not use the word *head* in connection with syntactic constructs to spare the linguist from ethical dilemmas. Just imagine what would happen if we replaced *head* by *leg* or *tentacle*.

In searching for better evidence of strict headedness, one should perhaps also take into account results from NLP and psycholinguistics. But it is beyond doubt that any reasonable evaluation of strict headedness presupposes a serious exploration of head-agnostic alternatives. This work is intended to be a first step in that direction.

ACKNOWLEDGMENTS

The idea for this article came from participating in a workshop on "Köpfigkeit und/oder grammatische Anarchie" ('Headedness and/or grammatical anarchy'), beautifully organized by Horst Simon and Ulrike Freywald, which took place at the Freie Universität in Berlin in May 2017. I am also very grateful to the reviewers of the Journal

[38] Let alone "monstres animaux" as formulated by Tesnière (1959, p. 347): "Les phrases bifides [e.g. coordination constructions] sont comparables aux monstres animaux, qui ont deux têtes ou deux extrémités inférieures." (Translation from Tesnière (2015, p. 351): "The bifid sentences are comparable to monsters that have two heads or two bodies.")

of Language Modelling for helpful comments & criticism, to the editors for their patience, and to Carmela Chateau-Smith who did a great job proof-reading the manuscript. Last but not least, the article has benefited enormously from discussions with Laura Kallmeyer, Rainer Osswald, Simon Petitjean, Agata Savary, and Christian Wurm. This work was partially financed by the CRC 991 "The Structure of Representations in Language, Cognition, and Science" at the University of Düsseldorf.

REFERENCES

Anne ABEILLÉ (2003), A Lexicon- and Construction-Based Approach to Coordinations, in Stefan MÜLLER, editor, *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar, Michigan State University*, pp. 5–25, CSLI Publications, Stanford, CA, `https://web.stanford.edu/group/cslipublications/cslipublications/HPSG/2003/abeille.pdf`.

Anne ABEILLÉ and Rui P. CHAVES (2021), Coordination, in Stefan MÜLLER, Anne ABEILLÉ, Robert D. BORSLEY, and Jean-Pierre KOENIG, editors, *Head-Driven Phrase Structure Grammar: The Handbook*, number 9 in Empirically Oriented Theoretical Morphology and Syntax, pp. 725–776, Language Science Press, Berlin, doi:10.5281/zenodo.5599848.

Anne ABEILLÉ and Owen RAMBOW (2000), Tree Adjoining Grammar: An Overview, in Anne ABEILLÉ and Owen RAMBOW, editors, *Tree Adjoining Grammars: Formalisms, Linguistic Analyses and Processing*, number 107 in CSLI Lecture Notes, pp. 1–68, CSLI Publications, Stanford, CA.

Anthony E. ADES and Mark J. STEEDMAN (1982), On the Order of Words, *Linguistics and Philosophy*, 4(4):517–558.

Vilmos ÁGEL (2000), *Valenztheorie*, Gunter Narr Verlag, Tübingen.

Kazimierz AJDUKIEWICZ (1935), Die syntaktische Konnexität, *Studia Philosophica*, 1:1–27.

Yehoshua BAR-HILLEL (1953), A Quasi-Arithmetical Notation for Syntactic Description, *Language*, 29(1):47–58.

Sascha BARGMANN (2015), Syntactically Flexible VP-Idioms and the N-after-N Construction, Poster presented at the 5th General Meeting, PARSEME, 23–24 September 2015, Iaşi, `https://typo.uni-konstanz.de/parseme/index.php/2-general/138-admitted-posters-iasi-23-24-september-2015`.

John BEAVERS (2003), More Heads and Less Categories: A New Look at Noun Phrase Structure, in Stefan MÜLLER, editor, *Proceedings of the 10th International Conference on Head-Driven Phrase Structure Grammar, Michigan State University*, pp. 47–67, CSLI Publications, Stanford, CA, `http://web.stanford.edu/group/cslipublications/cslipublications/HPSG/2003/beavers.pdf`.

Leonard BLOOMFIELD (1935), *Language*, George Allen & Unwin, London.

Joan BRESNAN, Ash ASUDEH, Ida TOIVONEN, and Stephen WECHSLER (2016), *Lexical-Functional Syntax*, Wiley Blackwell, Chichester, UK, 2nd edition.

Marie CANDITO, Guy PERRIER, Bruno GUILLAUME, Corentin RIBEYRE, Karën FORT, Djamé SEDDAH, and Eric De La CLERGERIE (2014), Deep Syntax Annotation of the Sequoia French Treebank, in Nicoletta Calzolari (Conference CHAIR), Khalid CHOUKRI, Thierry DECLERCK, Hrafn LOFTSSON, Bente MAEGAARD, Joseph MARIANI, Asuncion MORENO, Jan ODIJK, and Stelios PIPERIDIS, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, European Language Resources Association (ELRA), Reykjavik, `http://www.lrec-conf.org/proceedings/lrec2014/pdf/494_Paper.pdf`.

Joan CHEN-MAIN (2006), *On the Generation and Linearization of Multi-Dominance Structures*, Ph.D. thesis, Johns Hopkins University, Baltimore, MD.

Noam CHOMSKY (1957), *Syntactic Structures*, Mouton, Den Haag.

Noam CHOMSKY (1965), *Aspects of the Theory of Syntax*, MIT Press, Cambridge, MA.

Noam CHOMSKY (1970), Remarks on Nominalization, in Roderick JACOBS and Peter ROSENBAUM, editors, *Readings in English Transformational Grammar*, pp. 184–221, Ginn, Waltham, MA.

Noam CHOMSKY (1972), The Formal Nature of Language, in *Language and Mind*, pp. 115–160, Harcourt Brace Jovanovich, New York, 2nd edition.

Noam CHOMSKY (1981), *Lectures on Government and Binding*, Foris, Dordrecht.

Noam CHOMSKY (1986), *Barriers*, number 13 in Linguistic Inquiry Monographs, MIT Press, Cambridge, MA.

Noam CHOMSKY (2008), On Phases, in Robert FREIDIN, Carlos P. OTERO, and Maria Luisa ZUBIZARRETA, editors, *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, pp. 133–66, MIT Press, Cambridge, MA.

Noam CHOMSKY (2013), Problems of Projection, *Lingua*, 130:33–49.

William CROFT (1996), What's a Head?, in Johan ROORYCK and Laurie ZARING, editors, *Phrase Structure and the Lexicon*, pp. 35–75, Springer, Dordrecht, NL, doi:10.1007/978-94-015-8617-7_3.

Haskell B. CURRY (1961), Some Logical Aspects of Grammatical Structure, in Roman JAKOBSON, editor, *Structure of Language and its Mathematical Aspects: Proceedings of the Twelfth Symposium in Applied Mathematics held in New York City, April 14–15, 1960*, pp. 56–68.

Marie-Catherine DE MARNEFFE, Timothy DOZAT, Natalia SILVEIRA, Katri HAVERINEN, Filip GINTER, Joakim NIVRE, and Christopher D. MANNING (2014), Universal Stanford Dependencies: A Cross-linguistic Typology, in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pp. 4585–4592, European Language Resources Association (ELRA), Reykjavik, `http://www.lrec-conf.org/proceedings/lrec2014/pdf/1062_Paper.pdf`.

Ralph DEBUSMANN, Denys DUCHIER, and Geert-Jan M. KRUIJFF (2004), Extensible Dependency Grammar: A New Methodology, in Geert-Jan M. KRUIJFF and Denys DUCHIER, editors, *Proceedings of the COLING 2004 Workshop on Recent Advances in Dependency Grammar*, pp. 70–76, Geneva, `https://aclanthology.org/W04-1510`.

Robert FRANK (2002), *Phrase Structure Composition and Syntactic Dependencies*, MIT Press, Cambridge, MA.

Haim GAIFMAN (1965), Dependency Systems and Phrase-structure Systems, *Information and Control*, 8(3):304–337, doi:10.1016/S0019-9958(65)90232-9, also published as report p-2315, RAND Corp. Santa Monica (CA) in 1961.

Gerald GAZDAR (1981), Unbounded Dependencies and Coordinated Structure, *Linguistic Inquiry*, 12:155–184.

Gerald GAZDAR, Ewan KLEIN, Geoffrey K. PULLUM, and Ivan A. SAG (1985), *Generalized Phrase Structure Grammar*, Harvard University Press, Cambridge, MA.

Kim GERDES (2004), Tree Unification Grammar: Problems and Proposals for Topology, TAG, and German, *Electronic Notes in Theoretical Computer Science*, 53:108–136, doi:10.1016/S1571-0661(05)82578-2.

Kim GERDES and Sylvain KAHANE (2013), Defining Dependencies (and Constituents), in Kim GERDES, Eva HAJIČOVÁ, and Leo WANNER, editors, *Computational Dependency Theory*, number 258 in Frontiers in Artificial Intelligence and Applications, pp. 1–25, IOS Press, Amsterdam.

Jonathan GINZBURG and Ivan A. SAG (2001), *Interrogative Investigations: The Form, Meaning, and Use of English Interrogatives*, number 123 in CSLI Lecture Notes, CSLI Publications, Stanford, CA.

Maurice GROSS (1994), Constructing Lexicon-Grammars, in Beryl T. Sue ATKINS and Anotonio ZAMPOLLI, editors, *Computational Approaches to the Lexicon*, pp. 213–263, Oxford University Press, Oxford.

Bruno GUILLAUME and Guy PERRIER (2009), Interaction Grammars, *Research on Language and Computation*, 7(2–4):171–208.

Zellig S. HARRIS (1951), *Methods in Structural Linguistics*, The University of Chicago Press.

Peter HELLWIG (2006), Parsing with Dependency Grammars, in Vilmos ÁGEL, Ludwig M. EICHINGER, Hans-Werner EROMS, Hans Jürgen HERINGER, and Henning LOBIN, editors, *Dependenz und Valenz/Dependency and Valency. Ein internationales Handbuch der zeitgenössischen Forschung / An International Handbook of Contemporary Research*, volume 25.2 of *Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (HSK)*, pp. 1081–1109, Walter de Gruyter, Berlin.

Hans Jürgen HERINGER (1993), Basic Ideas and the Classical Model, in Joachim JACOBS, Arnim VON STECHOW, Wolfgang STERNEFELD, and Theo VENNEMANN, editors, *Syntax. Ein internationales Handbuch zeitgenössischer Forschung / An International Handbook of Contemporary Research*, volume 9.1 of *Handbücher zur Sprach- und Kommunikationswissenschaft / Handbooks of Linguistics and Communication Science (HSK)*, pp. 298–316, Walter de Gruyter, Berlin.

Jack HOEKSEMA (1992), The Head Parameter in Morphology and Syntax, *Language and Cognition*, 2:119–132.

Richard A. HUDSON (1984), *Word Grammar*, Blackwell, Oxford.

Richard A. HUDSON (1987), Zwicky on Heads, *Journal of Linguistics*, 23(1):109–132.

Richard A. HUDSON (1993), Do We Have Heads in Our Minds?, in Greville G. CORBETT, Norman M. FRASER, and Scott MCGLASHAN, editors, *Heads in Grammatical Theory*, pp. 266–291, Cambridge University Press, Cambridge, UK, doi:10.1017/CBO9780511659454.012.

Richard A. HUDSON (1994), Discontinuous Phrases in Dependency Grammar, *UCL Working Papers in Linguistics*, 6:89–120.

Richard A. HUDSON (1998), Adjunct Preposing, Wh-interrogatives and Dependency Competition, *UCL Working Papers in Linguistics*, 10:441–466.

Richard A. HUDSON (2004), Are Determiners Heads?, *Functions of Language*, 11(1):7–42.

Richard A. HUDSON (2007), *Language Networks: The New Word Grammar*, Oxford University Press, Oxford.

Ray S. JACKENDOFF (1977), *$\overline{X}$ Syntax: A Study of Phrase Structure*, MIT Press, Cambridge, MA.

Joachim JACOBS (1994), *Kontra Valenz*, Wissenschaftlicher Verlag Trier, Trier.

Otto JESPERSEN (1924), *The Philosophy of Grammar*, George Allen & Unwin, London.

Aravind K. Joshi and Yves Schabes (1997), Tree-Adjoining Grammars, in Grzegorz Rozenberg and Arto Salomaa, editors, *Handbook of Formal Languages*, volume 3, pp. 69–124, Springer, Berlin.

Aravind K. Joshi, K. Vijay Shanker, and David Weir (1990), The Convergence of Mildly Context-sensitive Grammar Formalisms, CIS Technical Report MS-CIS-90-01, Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, `http://repository.upenn.edu/cis_reports/539/`.

Sylvain Kahane (1997), Bubble Trees and Syntactic Representations, in Tilman Becker and Hans-Ulrich Krieger, editors, *Proceedings of the 5th Meeting on Mathematics of Language (MOL5)*, pp. 70–76, DFKI, Saarbrücken.

Sylvain Kahane (2006), Polarized Unification Grammars, in *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp. 137–144, Sydney, `http://www.aclweb.org/anthology/P/P06/P06-1018`.

Laura Kallmeyer and Marco Kuhlmann (2012), A Formal Model for Plausible Dependencies in Lexicalized Tree Adjoining Grammar, in Chung-Hye Han and Giorgio Satta, editors, *Proceedings of the 11th International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+11), 26–28 September 2012, Paris, France*, pp. 108–116, `https://aclanthology.org/W12-4613.pdf`.

Ron M. Kaplan and Joan Bresnan (1982), Lexical-Functional Grammar: A Formal System for Grammatical Representation, in Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*, pp. 173–281, MIT Press, Cambridge, MA.

Andreas Kathol (1995), *Linearization-Based German Syntax*, Doctoral thesis, Ohio State University, Columbus, OH.

András Kornai and Geoffrey K. Pullum (1990), The X-Bar Theory of Phrase Structure, *Language*, 66(1):24–50, `http://www.jstor.org/stable/415278`.

Paul R. Kroeger (2005), *Analyzing Grammar: An Introduction*, Cambridge University Press, Cambridge.

Marco Kuhlmann (2010), *Dependency Structures and Lexicalized Grammars: An Algebraic Approach*, number 6270 in Lecture Notes in Computer Science, Springer, Berlin.

Timm Lichte (2012), Synchronous Tree Unification Grammar, in Chung-Hye Han and Giorgio Satta, editors, *Proceedings of the 11th International Workshop on Tree Adjoining Grammar and Related Formalisms (TAG+11), 26–28 September 2012, Paris, France*, pp. 46–54, `http://www.aclweb.org/anthology-new/W/W12/W12-4606.pdf`.

Timm LICHTE (2015), *Syntax und Valenz: Zur Modellierung kohärenter und elliptischer Strukturen mit Baumadjunktionsgrammatiken*, number 1 in Empirically Oriented Theoretical Morphology and Syntax, Language Science Press, Berlin, doi:10.17169/langsci.b20.102.

Timm LICHTE and Laura KALLMEYER (2016), Same Syntax, Different Semantics: A Compositional Approach to Idiomaticity in Multi-word Expressions, in Christopher PIÑÓN, editor, *Empirical Issues in Syntax and Semantics 11*, pp. 111–140, `http://www.cssp.cnrs.fr/eiss11/eiss11_lichte-and-kallmeyer.pdf`.

Timm LICHTE and Laura KALLMEYER (2017), Tree-Adjoining Grammar: A Tree-based Constructionist Grammar Framework for Natural Language Understanding, in Luc STEELS and Jerome FELDMAN, editors, *The AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding, March 27–29, 2017, Stanford University (Technical Report SS-17-02)*, pp. 205–212, `http://aaai.org/ocs/index.php/SSS/SSS17/paper/view/15330/14536`.

Timm LICHTE, Simon PETITJEAN, Agata SAVARY, and Jakub WASZCZUK (2019), Lexical Encoding Formats for Multi-word Expressions: The Challenge of "Irregular" Regularities, in Yannick PARMENTIER and Jakub WASZCZUK, editors, *Representation and Parsing of Multiword Expressions: Current Trends*, number 1 in Phraseology and Multiword Expressions, pp. 1–33, Language Science Press, Berlin, doi:10.5281/zenodo.2579033.

John LYONS (1968), *Introduction to Theoretical Linguistics*, Cambridge University Press, Cambridge.

John LYONS (1977), *Semantics*, Cambridge University Press, Cambridge.

Wolfgang MAIER and Timm LICHTE (2011), Characterizing Discontinuity in Constituent Treebanks, in Philippe DE GROOTE, Markus EGG, and Laura KALLMEYER, editors, *Formal Grammar. 14th International Conference, FG 2009. Bordeaux, France, July 25-26, 2009. Revised Selected Papers*, number 5591 in Lecture Notes in Artificial Intelligence, pp. 167–182, Springer, Berlin.

James D. MCCAWLEY (1982), Parentheticals and Discontinuous Constituent Structure, *Linguistic Inquiry*, 13(1):91–106.

Igor A. MEL'ČUK (1988), *Dependency Syntax: Theory and Practice*, State University of New York Press, Albany.

Igor A. MEL'ČUK (2004), Actants in Semantics and Syntax II: Actants in Syntax, *Linguistics*, 42(2):247–291, doi:https://doi.org/10.1515/ling.2004.009.

Igor A. MEL'ČUK (2009), Dependency in Natural Language, in Alain POLGUÈRE and Igor A. MEL'ČUK, editors, *Dependency in Linguistic Description*, number 111 in Studies in Language Companion Series, pp. 1–110, John Benjamins, Amsterdam, doi:10.1075/slcs.111.03mel.

Olivier MICHALON, Corentin RIBEYRE, Marie CANDITO, and Alexis NASR (2016), Deeper Syntax for Better Semantic Parsing, in *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers, December 11–16, 2016, Osaka, Japan*, pp. 409–420, `https://aclanthology.org/C16-1040.pdf`.

Stefan MÜLLER (1999), An HPSG-Analysis for Free Relative Clauses in German, *Grammars*, 2:53–105, doi:10.1023/A:1004564801304.

Stefan MÜLLER (2013), Unifying Everything: Some Remarks on Simpler Syntax, Construction Grammar, Minimalism and HPSG, *Language*, 89(4):920–950, `https://hpsg.hu-berlin.de/~stefan/Pub/unifying-everything.html`.

Stefan MÜLLER (2013 [2007]), *Head-Driven Phrase Structure Grammar: Eine Einführung*, number 17 in Stauffenburg Einführungen, Stauffenburg Verlag, Tübingen, dritte überarbeitete Auflage edition, `http://hpsg.fu-berlin.de/~stefan/Pub/hpsg-lehrbuch.html`.

Stefan MÜLLER (2015), The CoreGram Project: Theoretical Linguistics, Theory Development, and Verification, *Journal of Language Modelling*, 3(1):21–86, doi:10.15398/jlm.v3i1.91.

Stefan MÜLLER (2020a), *Grammatical Theory: From Transformational Grammar to Constraint-Based Approaches*, number 1 in Textbooks in Language Sciences, Language Science Press, Berlin, fourth revised and extended edition edition, doi:10.5281/zenodo.3992307, `https://langsci-press.org/catalog/book/287`.

Stefan MÜLLER (2020b), Headless in Berlin: Headless Structures in Head-Driven Phrase Structure Grammar, in Ulrike FREYWALD and Horst SIMON, editors, *In Headedness and/or Grammatical Anarchy?*, pp. 1–32, Language Science Press, Berlin, `https://hpsg.hu-berlin.de/~stefan/Pub/headless.html`.

Stefan MÜLLER and Stephen M. WECHSLER (2014), Lexical Approaches to Argument Structure, *Theoretical Linguistics*, 40(1–2):1–76, `http://hpsg.fu-berlin.de/~stefan/Pub/arg-st.html`.

Joakim NIVRE (2015), Towards a Universal Grammar for Natural Language Processing, in Alexander GELBUKH, editor, *Computational Linguistics and Intelligent Text Processing: 16th International Conference, CICLing 2015, Cairo, Egypt, April 14–20, 2015, Proceedings, Part I*, pp. 3–16, Springer, Cham, doi:10.1007/978-3-319-18111-0_1.

Carl POLLARD and Ivan A. SAG (1994), *Head-Driven Phrase Structure Grammar*, University of Chicago Press, Chicago, IL.

Fred POPOWICH (1989), Tree Unification Grammar, in *27th Annual Meeting of the Association for Computational Linguistics*, pp. 228–236, Association for Computational Linguistics, Vancouver, British Columbia, Canada, doi:10.3115/981623.981651, `https://aclanthology.org/P89-1028.pdf`.

Adam PRZEPIÓRKOWSKI and Agnieszka PATEJUK (2020), From Lexical Functional Grammar to Enhanced Universal Dependencies: The UD-LFG Treebank of Polish, *Language Resources and Evaluation*, 54:185–221, doi:10.1007/s10579-018-9433-z.

Geoffrey K. PULLUM and Barbara C. SCHOLZ (2001), On the Distinction Between Model-theoretic and Generative-enumerative Syntactic Frameworks, in Philippe DE GROOTE, Glyn MORILL, and Christian RETORÉ, editors, *Logical Aspects of Computational Linguistics: 4th International Conference, LACL 2001 Le Croisic, France, June 27–29, 2001*, pp. 17–43, Springer, Berlin, doi:10.1007/3-540-48199-0_2.

Andrew RADFORD (2009), *An Introduction to English Sentence Structure*, Cambridge University Press, Cambridge.

Owen RAMBOW, K. VIJAY-SHANKER, and David WEIR (1995), D-Tree Grammars, in *Proceedings of the 33rd Annual Conference of the Association for Computational Linguistics*, pp. 151–158, Association for Computational Linguistics, Cambridge, MA, doi:10.3115/981658.981679, `https://www.aclweb.org/anthology/P95-1021.pdf`.

Anoop SARKAR and Aravind K. JOSHI (1997), Handling Coordination in a Tree Adjoining Grammar, `http://www.cs.sfu.ca/~anoop/papers/pdf/tag-coordination.pdf`, longer version of paper in Proceedings of COLING 1996. Draft of August 19, 1997.

Harold L. SOMERS (1984), On the Validity of the Complement-adjunct Distinction in Valency Grammar, *Linguistics*, (22):507–530, doi:10.1515/ling.1984.22.4.507.

Edward P. STABLER (2011), Computational Perspectives on Minimalism, in Cedric BOECKX, editor, *The Oxford Handbook of Linguistic Minimalism*, Oxford Handbooks in Linguistics, chapter 27, pp. 616–641, Oxford University Press, Oxford.

Mark STEEDMAN (2000), *The Syntactic Process*, MIT Press, Cambridge, MA.

Martin STOKHOF and Michiel VAN LAMBALGEN (2011), Abstractions and Idealisations: The Construction of Modern Linguistics, *Theoretical Linguistics*, 37:1–26, doi:10.1515/THLI.2011.001.

Angelika STORRER (1992), *Verbvalenz. Theoretische und methodische Grundlagen ihrer Beschreibung in Grammatikographie und Lexikographie*, number 126 in Reihe Germanistische Linguistik, Niemeyer, Tübingen.

Lucien TESNIÈRE (1959), *Eléments de syntaxe structurale*, Klincksieck, Paris.

Lucien TESNIÈRE (2015), *Elements of structural syntax*, John Benjamins, Amsterdam, translated by Timothy Osborne and Sylvain Kahane.

Willy Van Langendonck (1994), Determiners as Heads?, *Cognitive Linguistics*, 5(3):243–260, doi:10.1515/cogl.1994.5.3.243.

Willy van Langendonck (2003), The Dependency Concept and its Foundations, in Vilmos Ágel, Ludwig M. Eichinger, Hans-Werner Eroms, Hans Jürgen Heringer, and Henning Lobin, editors, *Dependenz und Valenz/Dependency and Valency. Ein internationales Handbuch der zeitgenössischen Forschung/An International Handbook of Contemporary Research*, volume 25.1 of *Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science (HSK)*, pp. 170–187, Walter de Gruyter, Berlin.

Heinz Vater (1978), Probleme der Verbvalenz, *Leuvense Bijdragen*, 67:267–308.

Rulon S. Wells (1947), Immediate Constituents, *Language*, 23:81–117.

XTAG Research Group (2001), A Lexicalized Tree Adjoining Grammar for English, Technical report, Institute for Research in Cognitive Science, University of Pennsylvania, Philadelphia, PA.

Arnold M. Zwicky (1985), Heads, *Journal of Linguistics*, 21:1–29, doi:10.1017/S0022226700010008.

Arnold M. Zwicky (1986), Concatenation and Liberation, in *Papers from the General Session at the Twenty-Second Regional Meeting of the Chicago Linguistic Society*, pp. 65–74, Chicago Linguistic Society, Chicago, IL.

Arnold M. Zwicky (1993), Heads, Bases and Functors, in Greville G. Corbett, Norman M. Fraser, and Scott McGlashan, editors, *Heads in Grammatical Theory*, pp. 292–315, Cambridge University Press, Cambridge, UK.

Timm Lichte
0000-0002-9202-8320
timm.lichte@uni-tuebingen.de

University of Tübingen, Germany

Timm Lichte (2021), *Against strict headedness in syntax*, Journal of Language Modelling, 9(2):291–348
https://dx.doi.org/10.15398/jlm.v9i2.249

This work is licensed under the *Creative Commons Attribution 4.0 Public License*.
http://creativecommons.org/licenses/by/4.0/