Parsing Indian English News Headlines

Samapika Roy

Department of Humanistic Studies
Indian Institute of Technology (BHU), Varanasi, India

Under the Supervision

Of

Dr. Sukhada

Anil Kumar Singh

Background & Motivation: Rationale

- Headlines compress grammar to deliver maximum meaning with minimal words.
- Linguistic parsers, trained on full canonical English, may misinterpret them
- Limited research exists on parsing such non-canonical or reduced registers.
- Quantitative comparison between the canonical register and the reduced register

Background & Motivation: Context

- Examines the syntactic and computational structure of news headlines.
- Treats headlines to be in the reduced register marked by reduction and creativity
- Understanding reduced syntax relevant for linguistic theory and NLP
- Term Indian English used to refrain from generalized claims
- However, it might potentially be generalizable

Research Aims

- To understand and analyze the the syntax of news headline
- How it differs from the full sentence canonical register
- Developing feature-value computational representation for NHs

Research Objectives

- Identify structural and morpho-syntactic features of reduced syntax
- Compare reduced structures with their canonical equivalents.
- Formulate a category-based description of these contrasts
- Develop a feature–value schema to encode them formally
- Using computational schema to quantitatively analyze the differences
  - Between the two registers

Research Questions

1. What are the morphosyntactic characteristics of news headlines
    - Reduced register
2. How do existing parsers interpret such reduced syntactic forms
3. How can reduced syntax be represented computationally
    - To understand the register differences
    - To possibly improve parser performance

Scope and Limitations

- Focus on English (Indian) headlines published in national newspapers
  - Excludes other reduced registers: Slides with bullet points (this one!)
  - Non-standard (reduced) registers: Social media, microblogging
- Restricted to morphosyntactic features
- Qualitative and quantitative comparative analysis

Benefits of the Research

- The language resources may be used for register comparisons
- Preparation of parallel data of registers
- Extracting morphosyntactic information from such data
- Correct labeling of grammatical categories of reduced registers
- Designing representations of morphosyntactic difference between registers
- For better Information Retrieval systems for non-canonical registers
- Machine Translation and other NLP applications

Structure of the Thesis

- Chapter 1: Introduction
- Chapter 2: Literature Review
- Chapter 3: Data Collection, Cleaning, and Analysis
- Chapter 4: Reduced Register as Used in NHs: Observations
- Chapter 5: Reduced-Canonical Registers: Computational Representation
- Chapter 6: Canonical-Reduced Registers: Findings and Concluding Comments

Research Road Map

- Data collection and cleaning
- Parsing and error analysis
- Linguistic analysis
- Transformation guidelines
- Computational feature-value representation creation
- Comparison of registers in terms of the created representation

Previous Works on News Headlines

- Linguistic Manipulation: An Analysis of How Attitudes are Displayed in News Reporting (Nordlund, 2003)
- Linguistic Analysis of Newspaper Discourse in Theory and Practice (Pajunen,2008)
- A Linguistic-Stylistic Analysis of Newspaper Reportage (Agu, 2015)
- Tense in News Headlines (Hameed, 2008)
- The uses of the present tense in headlines (Chovanec, 2008)
- Of headlines & headlinese: Towards distinctive linguistic and pragmatic genericity (Isani, 2011)
- A Brief Study on the Language of Newspaper Headlines Used in “The New Light of Myanmar” (Moe, 2014)

Previous Work on News Headlines:

- Automatic extraction of news values from headline text (Alicja Piotrkowicz et al.,2017)
- Emotion classification of news headlines using svm (Kirange and Deshmukh, 2012)
- Generating headline summary from a document set (Sarkar and Bandyopadhyay, 2005)
- Analysis of the Relation Between Stock Price Returns and Headline News Using Text Categorization.(Takahashi, 2007)
- Headline Evaluation Experiment Results (Zajic, 2004)
- Event-driven Headline Generation (Rui Sun et al., 2015)
- Headline Generation based on Statistical Translation (Banko et al., 2000)
- Using Thematic Information in Statistical Headline Generation (Wan, et al., 2003)

Research Gap

- Lack of linguistic characterization of English (Indian) NHs
- Study of parsers for reduced register structures and text
- Absence of formal, category-based mapping
  - Between canonical and reduced registers
- No feature–value schema capturing structural reduction
- Absence of syntactico-semantic model that integrates:
  - Linguistic and computational perspectives

Contributions

1. NHs corpus: A corpus of ~20,000 headlines of English (Indian) NHs collected
2. Linguistic analysis of the NHs data
    - Study of different structures of NHs and words compositions
3. Guideline creation for transformation of reduced structures of NHs
4. Categorical and feature-value representation for differences
5. Qualitative and quantitative comparison of registers

Methodology Overview

- Corpus-based study on reduced structures of news headlines
- Focus on syntactic reduction, structural analysis
- Analytical framework combining linguistic and computational methods

Data Collection and Cleaning

- Collection:
  - From online newspapers: The Hindu, Hindustan Times, and Times of India
  - Manual and automated scraping; filtered for relevance and completeness
  - Corpus size: ~20,000 NHs with more than 3 lakh word tokens
- Cleaning:
  - Multiple repetition of the same text- e.g., the word ‘section’, headlines
  - Unwanted text: e.g., names of places

Parser Output and Error Analysis

- Parsing with existing parsers : e.g. Stanford Parser
- Compared with grammatical structures of a language to analyse the errors

Ex. NH: Boat capsize toll touches 21

Constituency parse output:

```
(ROOT
  (NP
    (NP (NNP Boat))
    (NP (JJ capsize) (NN toll) (NNS touches))
    (NP (CD 21))))
```

Parsing Output and Error Analysis

- Noun tagged as adverbs:

Headline: Men cook up an experience to nibble at

```
(ROOT (S
        (VP (ADVP (RB Men)) (VB cook) (PRT (RP up))
        (NP (DT an) (NN experience))
        (S
                (VP (TO to) (VP (VB nibble) (PP (IN at)))))))
```

Parsing Output and Error Analysis

- **Verbs marked as adjectives:**

  Headline: Boat capsize toll touches 21

  Output:

```
(ROOT (NP (NP (NNP Boat))
        (NP (JJ capsize)
            (NN toll) (NNS touches))
        (NP (CD 21))))
```

Common Parsing Errors Observed

- POS misclassification:
  - Proper nouns as verbs: Delhi rains floods roads
- Omitted determiners and auxiliaries confuse parser expectations
- Wrong attachment in phrases due to compact syntax
- Coordination ambiguity and subject drop issues

Linguistic Analysis

- Corpus details: Top 3 English newspapers in India
- Audit Bureau of Circulations, compiled by Media Research Users Council (MRUC)
- Indian Readership Survey (IRS) 2017
- Domain: General

Newspapers	The Hindu (TH)	Times of India (TOI)	Hindustan Times (HT)
Corpus (2016-17)	1,000	1,000	1,000
Corpus (2019-20)	1,000	1,000	1,000

Linguistic Analysis Dimensions

- Morphological analysis: compounding, clipping, and nominalization
- Syntactic analysis: ellipsis, omission, inversion
- Semantic analysis: idiomatic usage, wordplay, and cultural terms
- Register focus: reduced syntax as an adaptive strategy

Linguistic Analysis of Reduced Structures

- Declarative Headlines
    - Statements that relay information, adhere to basic SOV

Examples:

TH: Medical services in Mysuru likely to be hit today

HT: China isolated on Jammu and Kashmir in informal UNSC talks

TOI: Kuldeep Singh Rathore named as chief of Himachal Congress

Linguistic Analysis of Reduced Structures

- Interrogative:
  - Type 1: Simple Interrogatives:

    Examples:

      TH- Doctors' protest: Will govt. give in on contentious provisions of KPME Bill?

      HT: Will Maharashtra Rera's SRO filter benefit homebuyers eventually?

      TOI: What happens to Rishabh Pant now?

  - Type 2: Echo questions: Statements, do not involve WH-movement

    Examples:

      TH: Ranbir Kapoor plays a DJ in Brahmastra? An insider spills the beans

      HT- Spielberg's stand cost Michael Douglas Cannes Glory?

      TOI- Maharashtra to bail out 11,000 staffers with fake caste certificates?

Linguistic Analysis of Reduced Structures

- Imperative:

  Examples:

  Develop scientific temper

  Focus on environment

- Exclamative:

  Examples:

  Just for the health of it!

  Catch them young!

Linguistic Analysis of Reduced Structures

2. Types of Tenses:

- Present

Historical Present (Chovanec, 2008)	● UK court clears extradition of Dawood's aide Jabir Moti to US ● Youth dies at police station
Present Continuous	● Pinarayi protecting encroachers ● Rs. 330 cr. towards MNREGA wage payment pending

Linguistic Analysis of Reduced Structures

- Past

Simple Past	● Woman molested by beauty salon staffer ● Two killed in gaur attacks
Past Participle	● 11 bitten by dogs in Kollam ● Sandalwood trees stolen from C.V. Raman's home

Linguistic Analysis of Reduced Structures

- Future Time

Future (through infinitive)	State to commission survey on bonded labour Statute Bench to examine plea against M.M. Mani
Using modal verbs	First Rafale will land in India by 2019: Trappier 'Congress will play it fair'
Within a day (Glassman, 2015)	Art workshop concludes today Legislature session begins in Belagavi today

Linguistic Analysis of Reduced Structures

- Phrasal Verbs (Garnier & Schmitt, 2015; Liu & Myers, 2018)

Phrasal Verbs Type 1: Intransitive PVs.	Men cook up an experience to nibble at 150 volunteers clean up Hampankatta area
Phrasal Verbs Type 2: Transitive PVs.	Check-dams coming up across T.N. at a cost of Rs. 1,000 cr. Siddha council takes up vitiligo cause
Phrasal Prepositional Verbs	Centre for Defence Studies to be set up at Andhra University

Linguistic Analysis of Reduced Structures

- Word-Formation Processes:

Cliticization	Delhi's pollution levels rise again
Clipping (Apocope)	RBI relaxes 26% cap for ARCs
Abbreviation	Car stolen from Gzb man found in vacant plot
Acronym	MP Cong gen secy dies of Covid-19 after wife

Linguistic Analysis of Reduced Structures

- Circumstantial Compounding (Morpho-Syntax: Cui et al., 2018):
    - Coordinative compounds:
        - Two constituents generally of the same syntactic category
        - Which bear equal semantic weights.
            - Tube panic: 2 wanted for questioning
            - City beauty floors them
    - Subordinate compound:
        - Where first constituent, is the modifier
        - Modifies the second constituent, which is the head
            - Displacement fear grips tribals
            - Bangladesh promises help to arson victims

Linguistic Analysis of Reduced Structures

Code-mixing	One 'magarmach' down: Badal after Sajjan Kumar sent to jail
Scare-Quotes	Road map to regaining 'cleanest' tag
Quotes Without Speaker	'Aggression, violence are a reality of the world we live in today'

Linguistic Analysis of Reduced Structures

- Punctuations

Comma	Type 1: Speech-Speaker ● Need an NIA unit in Mangaluru, says BSY
	Type 2: To conjoin two incidents ● Amazon to create a million jobs, Goyal takes back criticism
	Type 3: Coordinate Conjunction ● Anurag Sharma thanks policemen, officers
	Type 4: Cause-effect ● Bhalswa landfill fires, smog have residents in chokehold

Linguistic Analysis of Reduced Structures

- Punctuation

Colon	Type 1: Speech-speaker ● Recruitment policy in T.N. flawed: TVK
	Type 2: Cause-effect ● Misuse of funds: Official held
	Type 3: Topic-Information ● Businessman murder case: no arrests yet
	Type 4: Topic- Description ● ADHD: The attention question ● Claridge's: The Cookbook

Linguistic Analysis of Reduced Structures

- Punctuation:

Semi-Colon	Type 1: Incident-result relation ● Drunk student rams auto-rickshaws in Chennai; one person killed
	Type 2: Two incidents ● 1,969 fishermen traced; search on for another 855
Ellipses	● For today's engagement... ● Let the games begin...

Linguistic Analysis of Reduced Structures

- Linguistic Items Dropped:

Dropping of the subject	Living up to a cinematic tradition Taking a trip down memory lane
Dropping of the verb	Charges against CJI Misra scurrilous
Conditional Dropping of Auxiliary	Church reformation celebrated Gujarat govt. is most corrupt: Rahul
Conditional Dropping of Article Same conditions for definite and indefinite articles	Man dies of injuries A blow to foes, saysEPS A celebration of Childhood

Linguistic Analysis of Reduced Structures

Other linguistics items observed:

- Demonstratives: That, this, these, those
- Quantifiers: Some, every, each
- Complementizers: That, if, whether, for
- Cardinal Noun Phrases- Ex. 11 bitten by dogs in Kollam

Linguistic Analysis of Reduced Structures

- Rhetorical devices used

Personification	Facebook says technical error caused vulgar translation of Xi Jinping's name
Metaphors	It is raining groundnuts
Ambiguity	The lob is here to stay
Puns with homophones	Hiding in ‘Plane’ sight.

Categorical Representation

- We classified them into the following broad categories:
  - Declarative
  - Historical Present
  - Echo Questions
  - Interrogative
  - Non-interrogative Wh
  - Aux Drop
  - NP Drop
  - Quotes without speakers
  - Punctuations
  - Fragments

Annotation: Headlines Classification

- Task: Students given headlines and provided with categories
- Map the headlines as per the best fitting categories
- Annotators
  - From various backgrounds were included
- Multiple headlines were given from each broad category
- Annotation Agreement Result
  - Cohen’s kappa: Fair agreement

Feature-value Representation

- Created a syntactico-semantic feature representation
  - Based on linguistic analysis on NHs corpus
- Headline_Structure: Single-line or multi-line
- Headline_Type: Fragment or Non-fragment
- Fragment_Type: Complex compounds, phrases, or subordinate clauses
- Non-Fragment_Type: Declarative, Imperative, interrogative, exclamatory

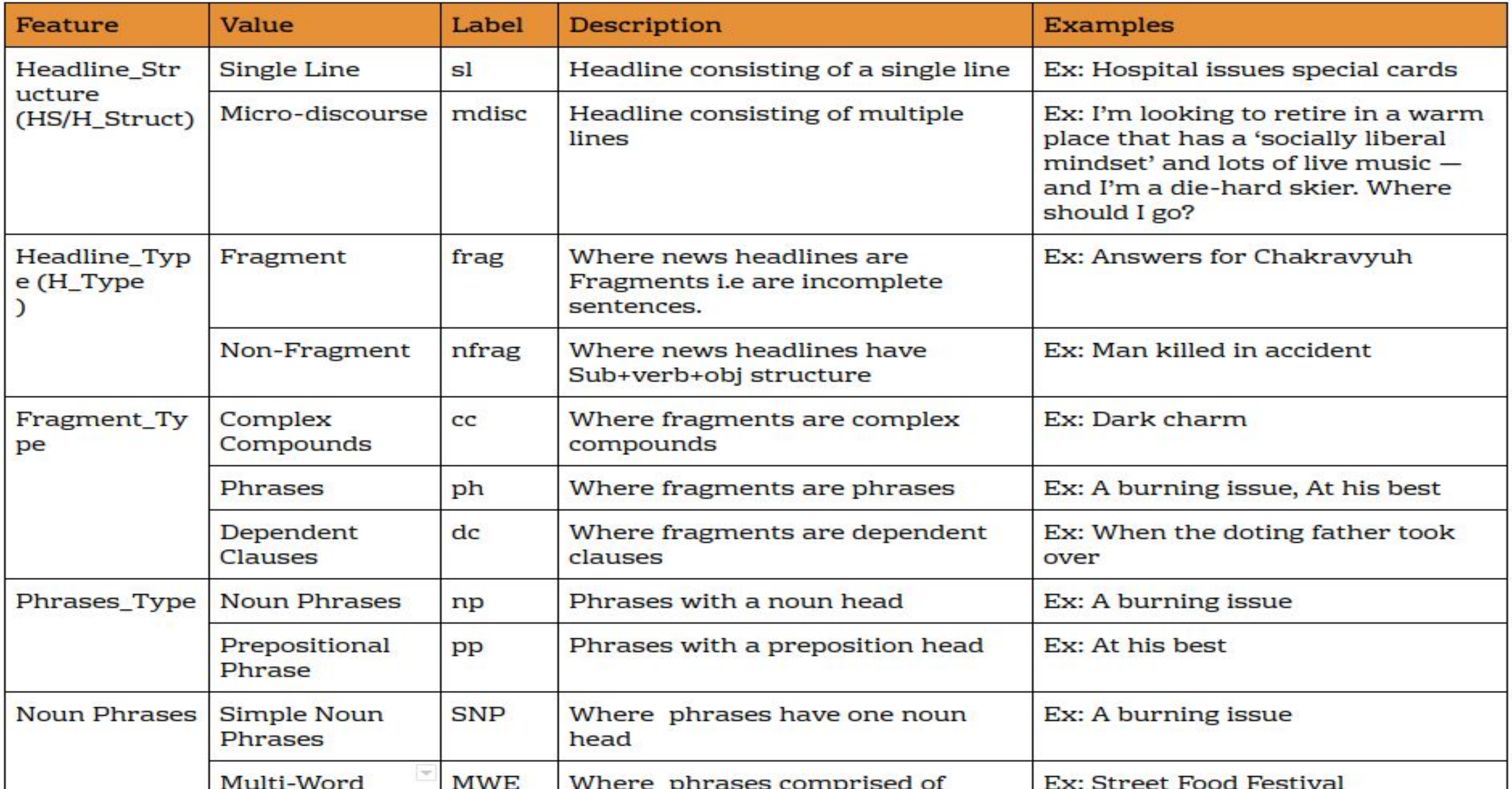
Structure of the Feature–Value Schema

- Features derived from linguistic categories:
  - Morphological
  - Tense, aspect, number
  - Syntactic: Clause type, subject presence, auxiliary omission
  - Functional: Focus, emphasis, information load
  - Each headline represented as a set of feature–value pairs

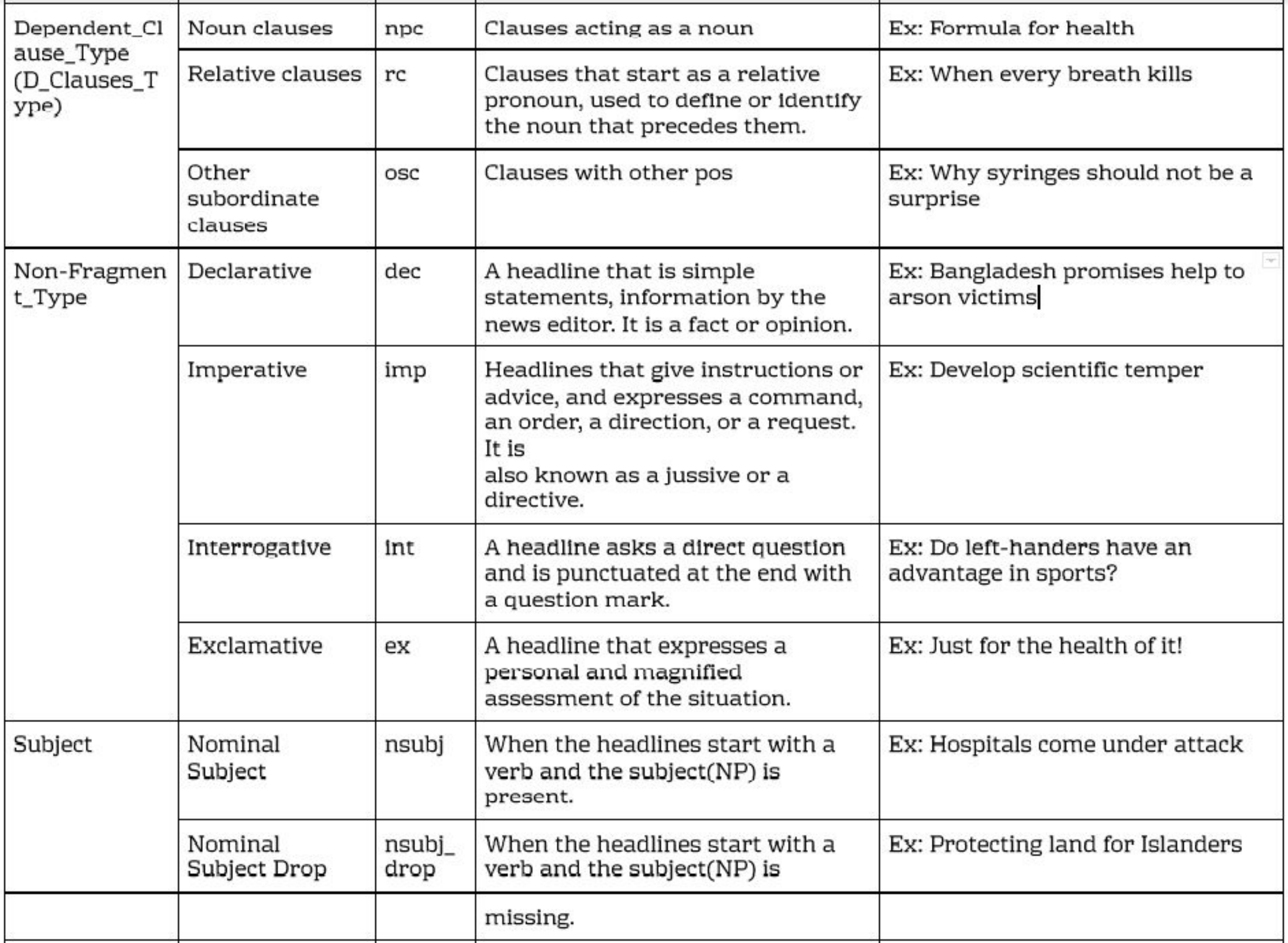
Reduced-Canonical Registers: Computational Representation

- Represents syntactic and morphological reductions
  - In terms of feature–value pairs
- Represent structural reduction
  - Through binary and categorical features
- Captures information beyond surface structure
- Allows formal encoding of reduced register grammar
  - For linguistic as well as computational purposes

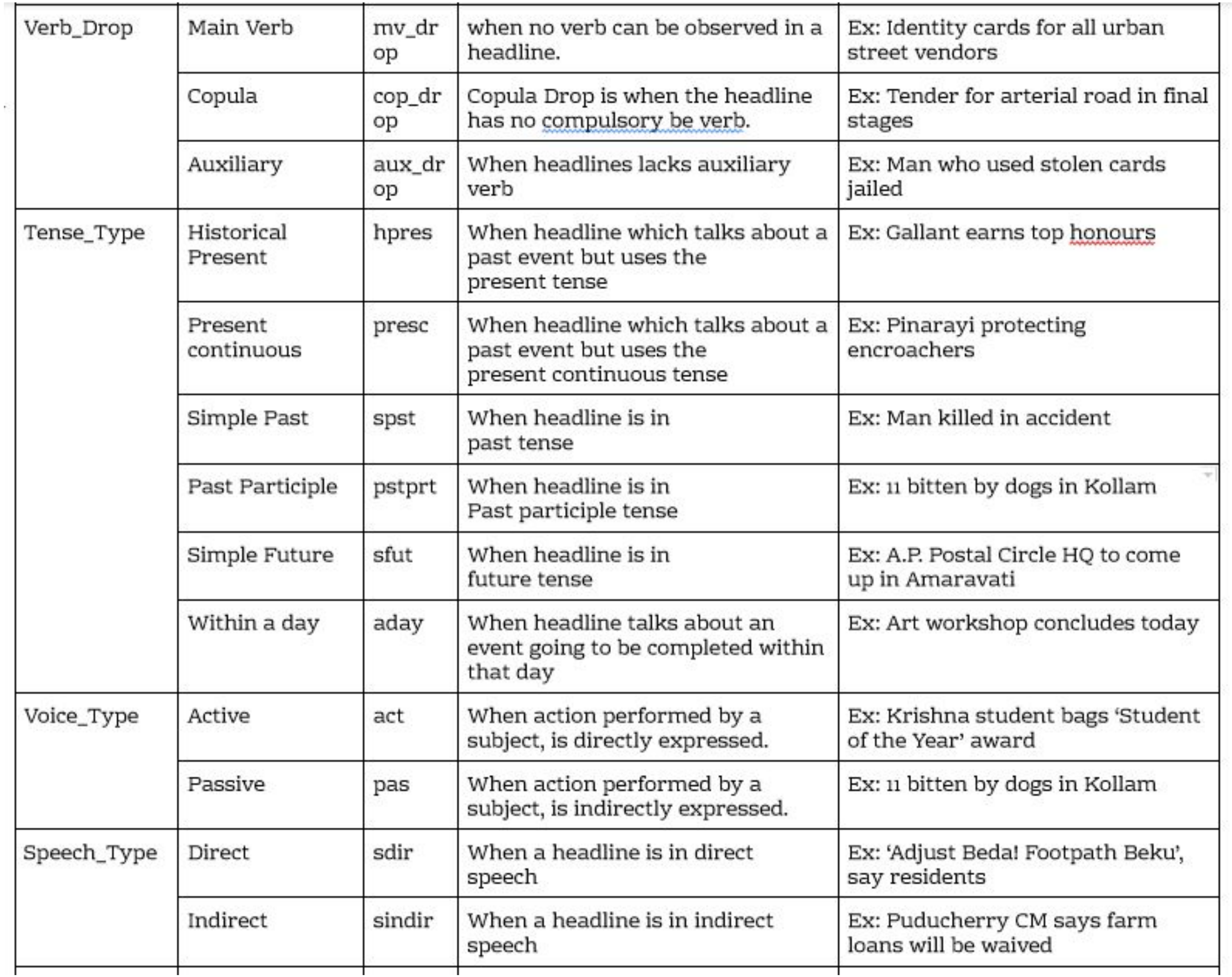
Feature-Value Model Phase II



Feature	Value	Label	Description	Examples
Headline_Structure (HS/H_Struct)	Single Line	sl	Headline consisting of a single line	Ex: Hospital issues special cards
	Micro-discourse	mdisc	Headline consisting of multiple lines	Ex: I'm looking to retire in a warm place that has a 'socially liberal mindset' and lots of live music — and I'm a die-hard skier. Where should I go?
Headline_Type (H_Type)	Fragment	frag	Where news headlines are Fragments i.e are incomplete sentences.	Ex: Answers for Chakravyuh
	Non-Fragment	nfrag	Where news headlines have Sub+verb+obj structure	Ex: Man killed in accident
Fragment_Type	Complex Compounds	cc	Where fragments are complex compounds	Ex: Dark charm
	Phrases	ph	Where fragments are phrases	Ex: A burning issue, At his best
	Dependent Clauses	dc	Where fragments are dependent clauses	Ex: When the doting father took over
Phrases_Type	Noun Phrases	np	Phrases with a noun head	Ex: A burning issue
	Prepositional Phrase	pp	Phrases with a preposition head	Ex: At his best
Noun Phrases	Simple Noun Phrases	SNP	Where phrases have one noun head	Ex: A burning issue
	Multi-Word	MWE	Where phrases comprised of	Ex: Street Food Festival



Dependent_Clause_Type (D_Clauses_Type)	Noun clauses	npc	Clauses acting as a noun	Ex: Formula for health
	Relative clauses	rc	Clauses that start as a relative pronoun, used to define or identify the noun that precedes them.	Ex: When every breath kills
	Other subordinate clauses	osc	Clauses with other pos	Ex: Why syringes should not be a surprise
Non-Fragment_Type	Declarative	dec	A headline that is simple statements, information by the news editor. It is a fact or opinion.	Ex: Bangladesh promises help to arson victims
	Imperative	imp	Headlines that give instructions or advice, and expresses a command, an order, a direction, or a request. It is also known as a jussive or a directive.	Ex: Develop scientific temper
	Interrogative	int	A headline asks a direct question and is punctuated at the end with a question mark.	Ex: Do left-handers have an advantage in sports?
	Exclamative	ex	A headline that expresses a personal and magnified assessment of the situation.	Ex: Just for the health of it!
Subject	Nominal Subject	nsubj	When the headlines start with a verb and the subject(NP) is present.	Ex: Hospitals come under attack
	Nominal Subject Drop	nsubj_drop	When the headlines start with a verb and the subject(NP) is missing.	Ex: Protecting land for Islanders



Verb_Drop	Main Verb	mv_drop	when no verb can be observed in a headline.	Ex: Identity cards for all urban street vendors
	Copula	cop_drop	Copula Drop is when the headline has no compulsory be verb.	Ex: Tender for arterial road in final stages
	Auxiliary	aux_drop	When headlines lacks auxiliary verb	Ex: Man who used stolen cards jailed
Tense_Type	Historical Present	hpres	When headline which talks about a past event but uses the present tense	Ex: Gallant earns top honours
	Present continuous	presc	When headline which talks about a past event but uses the present continuous tense	Ex: Pinarayi protecting encroachers
	Simple Past	spst	When headline is in past tense	Ex: Man killed in accident
	Past Participle	pstprt	When headline is in Past participle tense	Ex: 11 bitten by dogs in Kollam
	Simple Future	sfut	When headline is in future tense	Ex: A.P. Postal Circle HQ to come up in Amaravati
	Within a day	aday	When headline talks about an event going to be completed within that day	Ex: Art workshop concludes today
Voice_Type	Active	act	When action performed by a subject, is directly expressed.	Ex: Krishna student bags ‘Student of the Year’ award
	Passive	pas	When action performed by a subject, is indirectly expressed.	Ex: 11 bitten by dogs in Kollam
Speech_Type	Direct	sdir	When a headline is in direct speech	Ex: ‘Adjust Beda! Footpath Beku’, say residents
	Indirect	sindir	When a headline is in indirect speech	Ex: Puducherry CM says farm loans will be waived

Punctuation_Type (Punct_Type)	Comma	com	If Headline consists of a comma
	Colon	col	If Headline consists of a colon
	Semi-colon	scol	If Headline consists of a semi-colon

Annotation: Phase II

- Annotators given data to annotate
  - Covering different headline constructions
  - Detailed information about categories
  - same number of headlines covering various differences
- The annotators were linguists
- Cohen’s kappa: 80.65%
  - Indicating still some limitations remained

Feature-Value Annotation Model

Pattern:
For fragments:
H_Structure;H_Type; Fragment_Type;Phrase_Type;D_Clauses_Type

For Non-fragments:
H_Structure;H_Type;Non-Fragment_Type;subj_drop;V_Drop;Tense_type;Voice_Type;Speech_Type;Question_Type ;Punct_Type

Examples:

- Fragment:
  - Vegetable Dip: sl;frag;ph;cc
  - Women on top: sl;frag;ph;snp
- Non-Fragment:
  - 11 bitten by dog in Kollam: sl;nfrag;dec;nsubj;aux_drop;pstprt;pas;s_indir;0;0
  - Ryot killed by cow vigilantes, says family: sl;nfrag;dec;nsubj;aux_drop;spst;pas;s_indir;0;com

Comparison

Feature-value Phase I

- Flat categorization based on manual annotation
- Had only 10 broadly classified categories and no sub-categories
- Simplified

Feature-value Phase II

- Feature–value hierarchy allowing nested relations
- Have 13 Categories and sub-categories
- Detailed

Guideline Creation for Canonical Construction of NHs

- Guideline creation:
    - Covering the various structures of NHs found
    - From linguistic analysis
- Condition: If headlines are in present tense
    - Type 1: Singular verbs
- Solution:
    - Change verb to either present continuous or past construction
    - Depending on the occurrence of event

Guidelines Contd.

- Type 1: Singular verbs
- Type 1.a.: Regular verbs:
    - Art workshop concludes today
    - Art workshop is concluding today
- Type 1.b.: Irregular verbs:
    - Woman passenger falls to death
    - A Woman passenger fell to death

Guideline Contd.

- Type 2: Plural verbs:
  - Solution: Change verb to past tense for past events
- Type 2.a.: Plural_regular verbs
  - ICC nod for independent director
  - ICC noded for independent director
- Type 2.b.: Plural_irregular verbs
  - Hospitals come under attack
  - Hospitals came under attack

Parallel Corpus

- Parallel corpus of 4,000 (approx.) NHs
- Grammatically transformed canonical forms
- Manually aligned and validated
- Enables supervised training and evaluation of parser adaptation
- Examples:
    - Past:
        - Church reformation celebrated
        - The church reformation was celebrated
    - Present:
        - Rajini app crosses 1 lakh downloads
        - The Rajini app crossed 1 lakh downloads

Another Look at the Parallel Corpus

- On closer inspection, we found:
  - Converted ‘canonical’ forms were not really canonical
  - Missing some elements
    - Not full sentences
    - In almost all cases
- First goal of then
  - To make them really canonical
  - As full sentences
  - Wherever enough context for morphosyntactic interpretation possible
- Otherwise
  - Leave them as they are
  - So they don’t adversely affect the last and most important part
  - Of this work

Conversion to Truly Canonical Forms

- To avoid delay, tried to use LLMs
    - Iteratively refined instructions as prompts
    - Using multiple LLMs: Free versions from browsers
    - Prompt, input and output limits caused problems
    - Instructions had to be repeatedly given
- Many problems in the process
- Ultimately able to convert after several iterations
    - Fast manual inspection after each batch and every iteration
- Took almost as long as manual conversion would have

Further Checking of Alignment of Converted Batches

- Limitations of batch-wise conversion and multiple iterations
  - Ensuring alignment took time and some manual work
- Final checking of converted forms
- Manual correction in some cases
- At the end, truly parallel corpus
  - Reduced (NH) versions to canonical versions
- Used this very good quality resource as the base for further work

Starting with Created Feature-Value Representation

- Using manually created feature-value representation
    - Of register differences
- Converted to JSON schema
- Further extensive refinement of the schema
- Based on the very good quality register-parallel corpus
- Used the much better Stanza parser
    - Does both constituency and dependency parsing
    - Both kinds of parses created for the complete corpus
    - Reduced versions and canonical versions

Feature Extraction

- Using the parsed (constituency and dependency)
- Features extracted by a Python project created for this purpose
- Quantitative differences in terms of refined feature-value schema
- Tabular data about the register differences
- Most importantly
    - Instead of thinking in terms of parser ‘errors’
    - We focused on the register differences
    - Reason: The ‘errors’ are mostly register differences
    - There may be genuine errors too, obviously
    - But it is not fruitful to think in terms of these

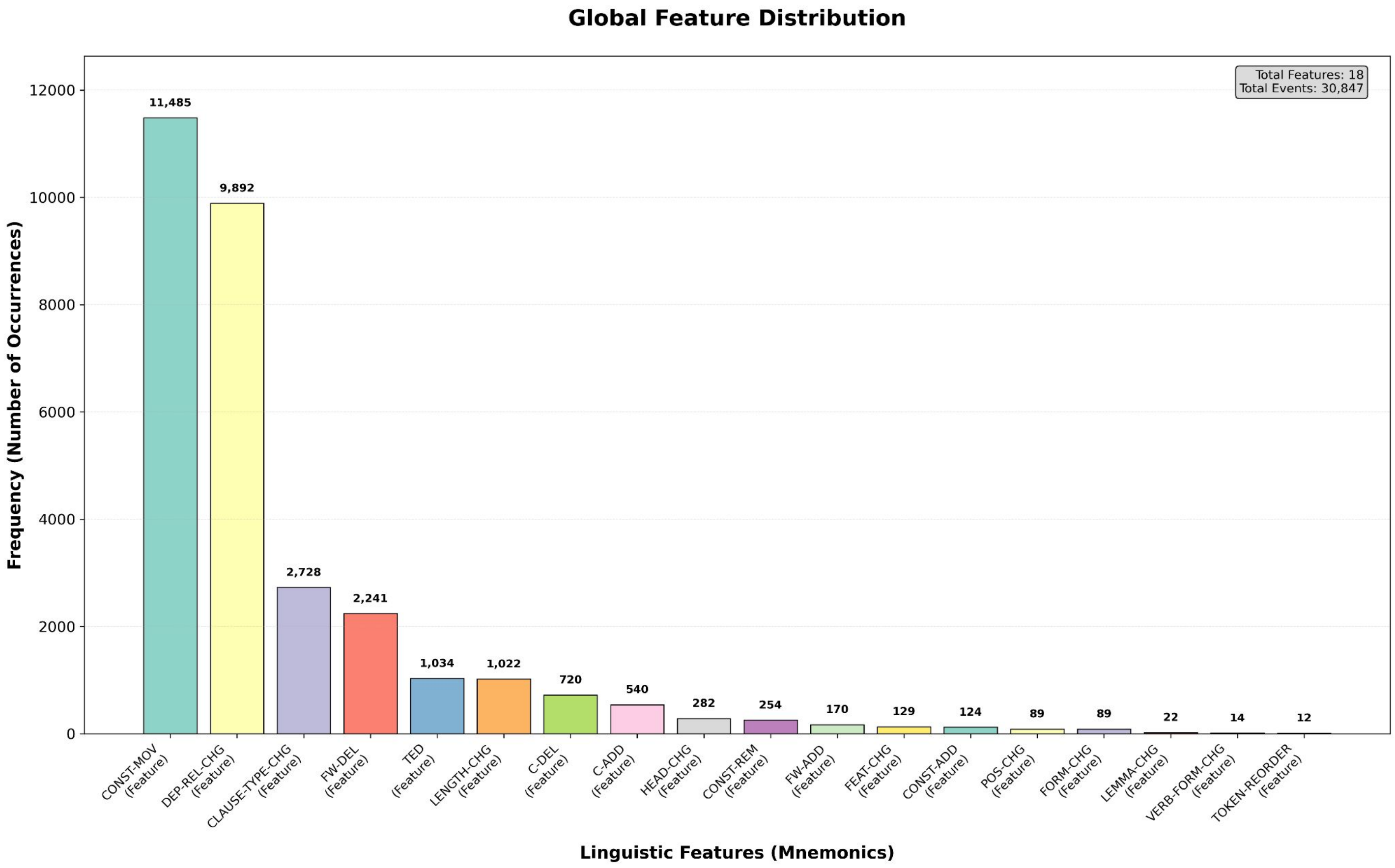
Further Refinement of the Schema

- Using the results obtained from the previous step
- Kinds of missing differences extracted
- Used to further refine the schema
- Several iterations
- But much faster this time using custom Python code

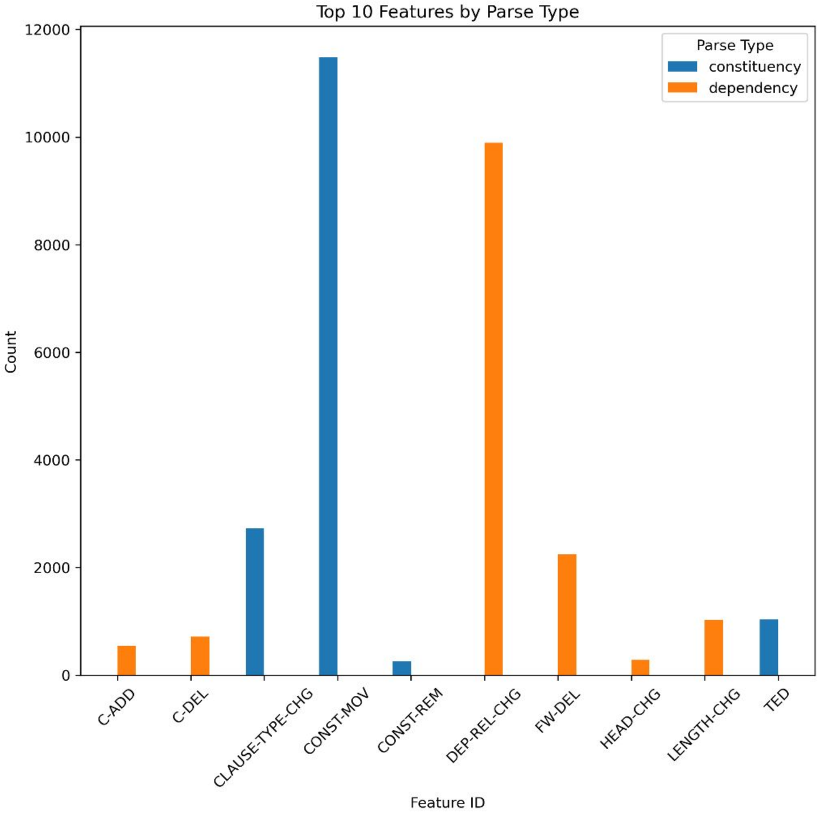
Visualization of Results

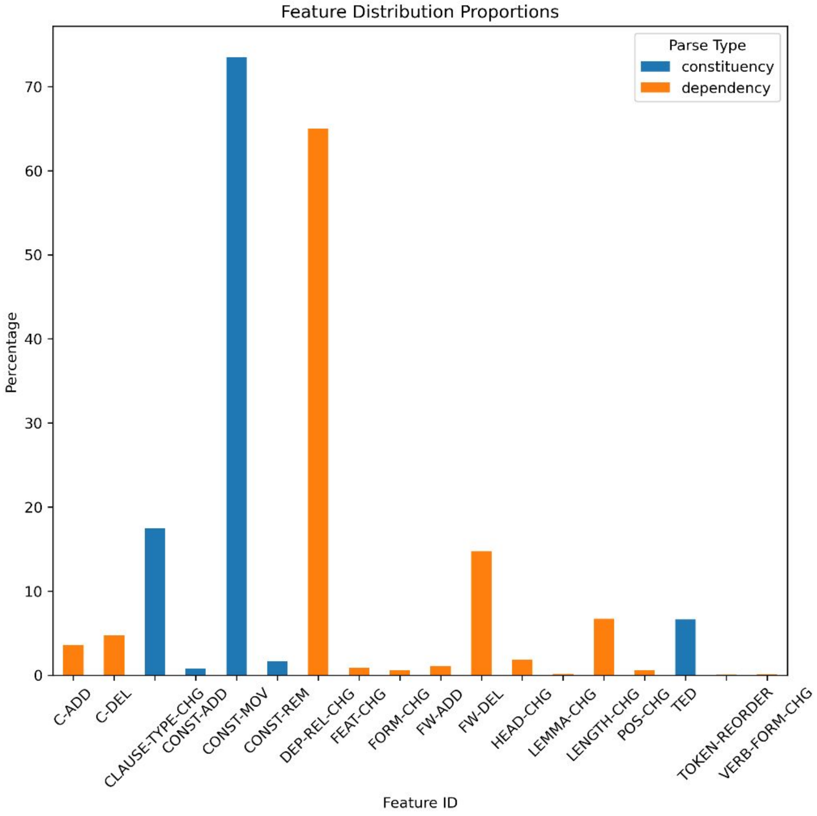
- Finally, using the refined schema and extracted quantitative differences
- Number of visualizations created
- Purpose being to get analytical insights about the register differences
- Could validate/contradict some/all of theoretical observations
  - In past work and in this work

Global Feature Distribution

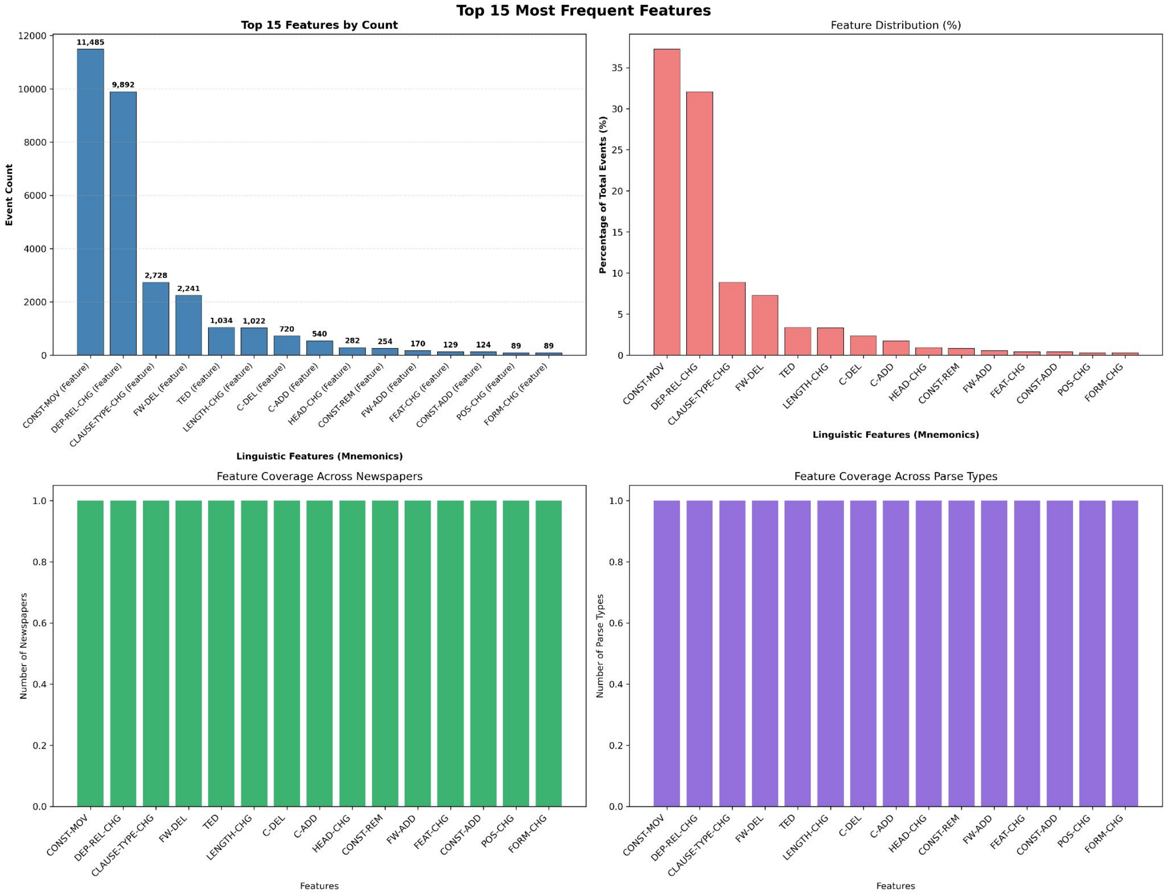


Parse-Type Comparison: All Feature-Value Pairs

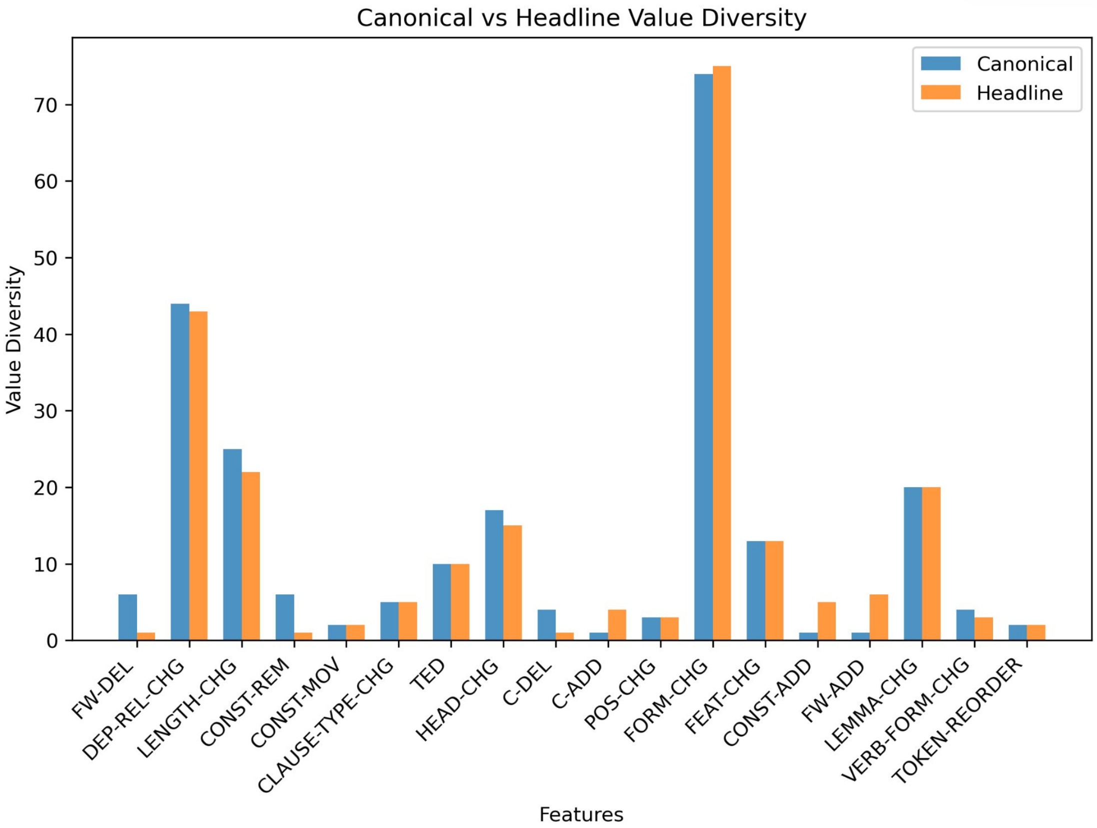




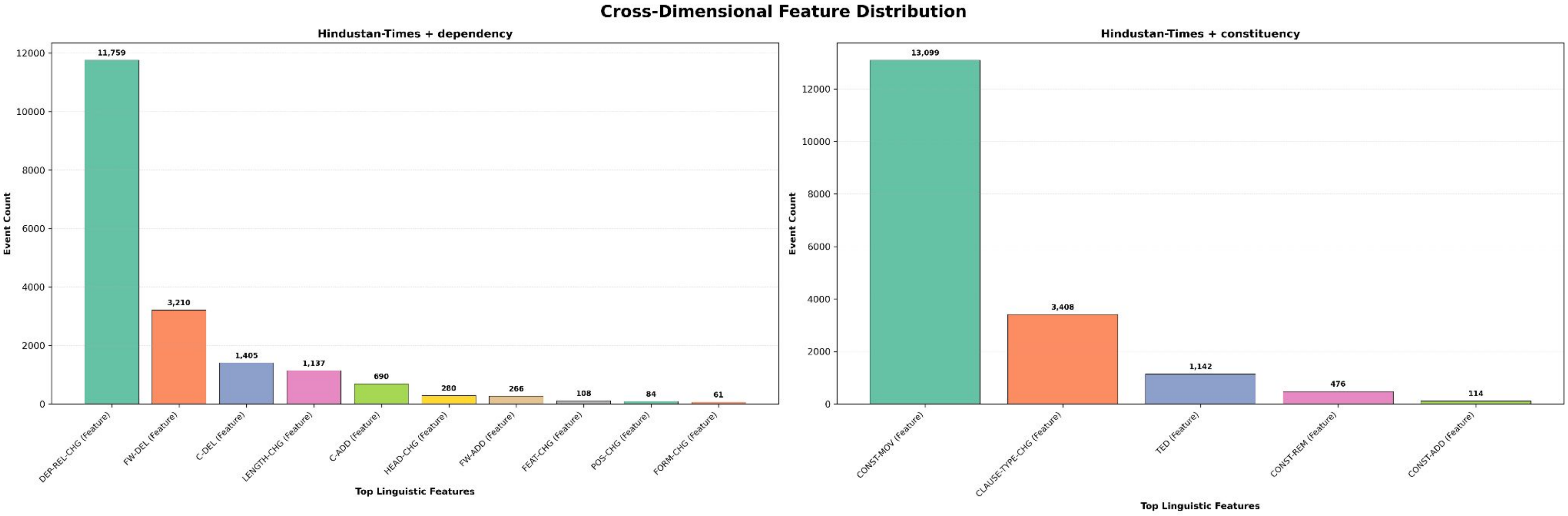
Top Feature-Value Transformations: All Pairs



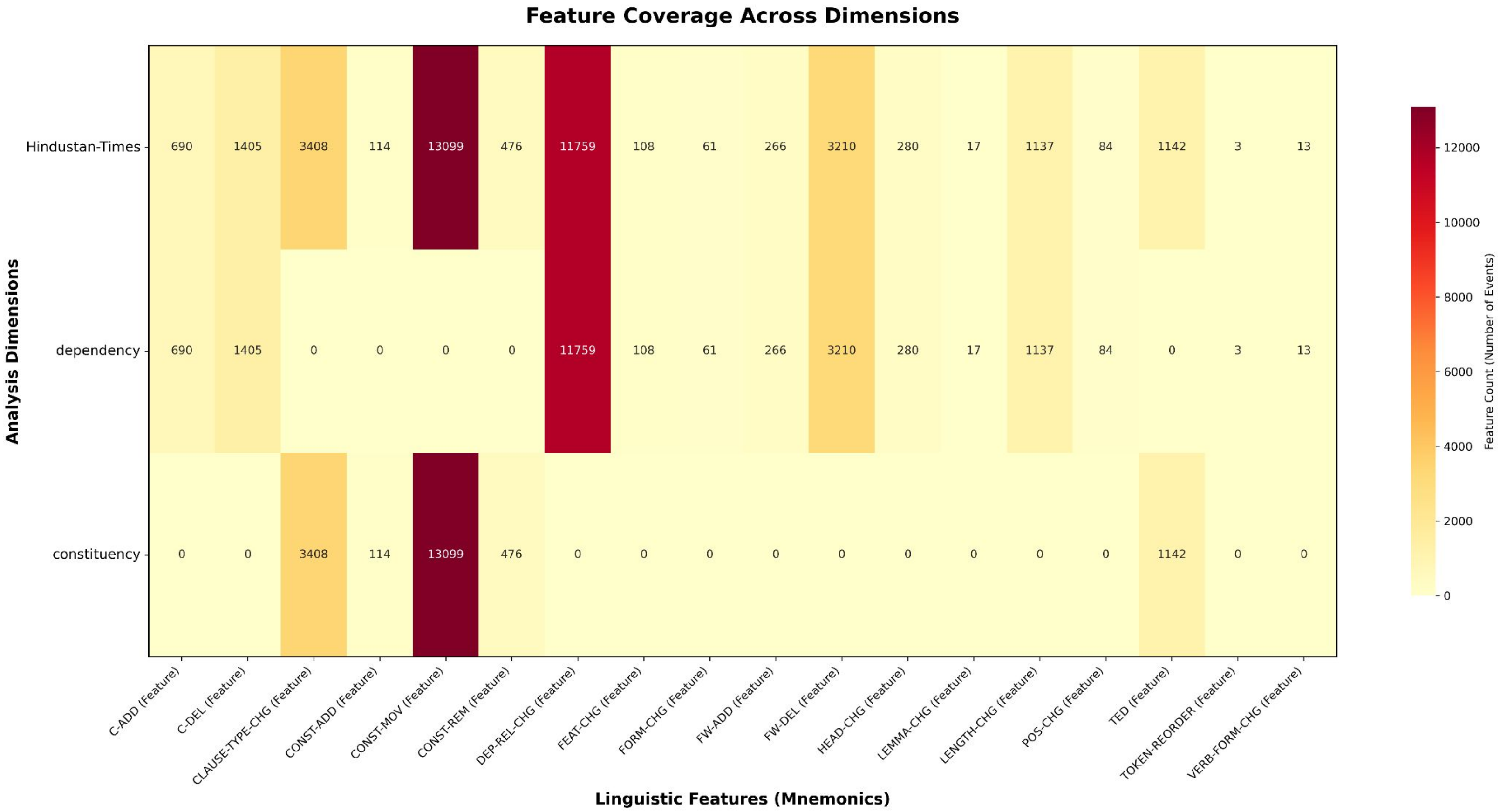
Canonical vs. Reduced: Diversity



Cross-Dimensional Feature-Value Distribution: HT

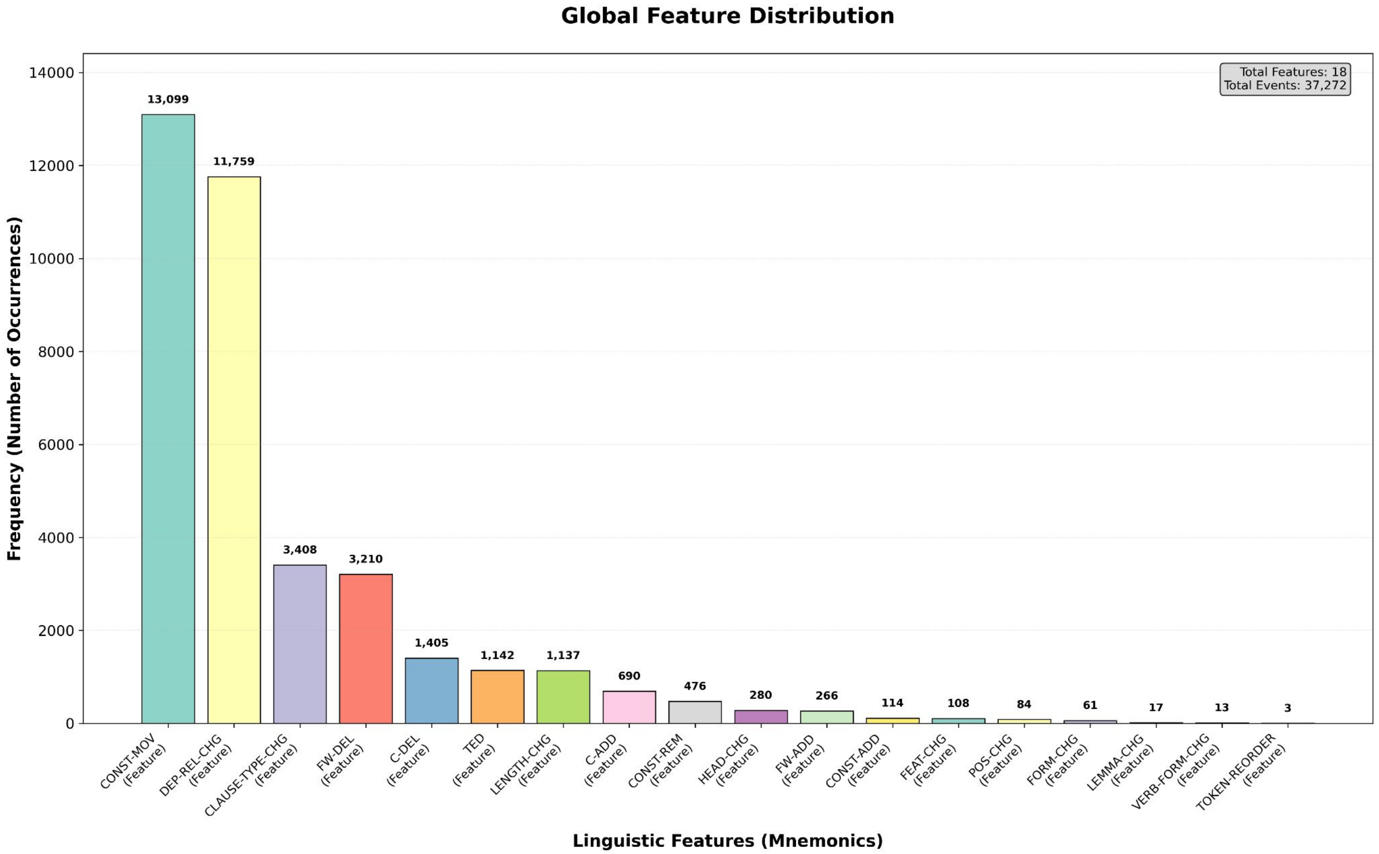


Feature-Value Pair Coverage Across Dimensions: HT

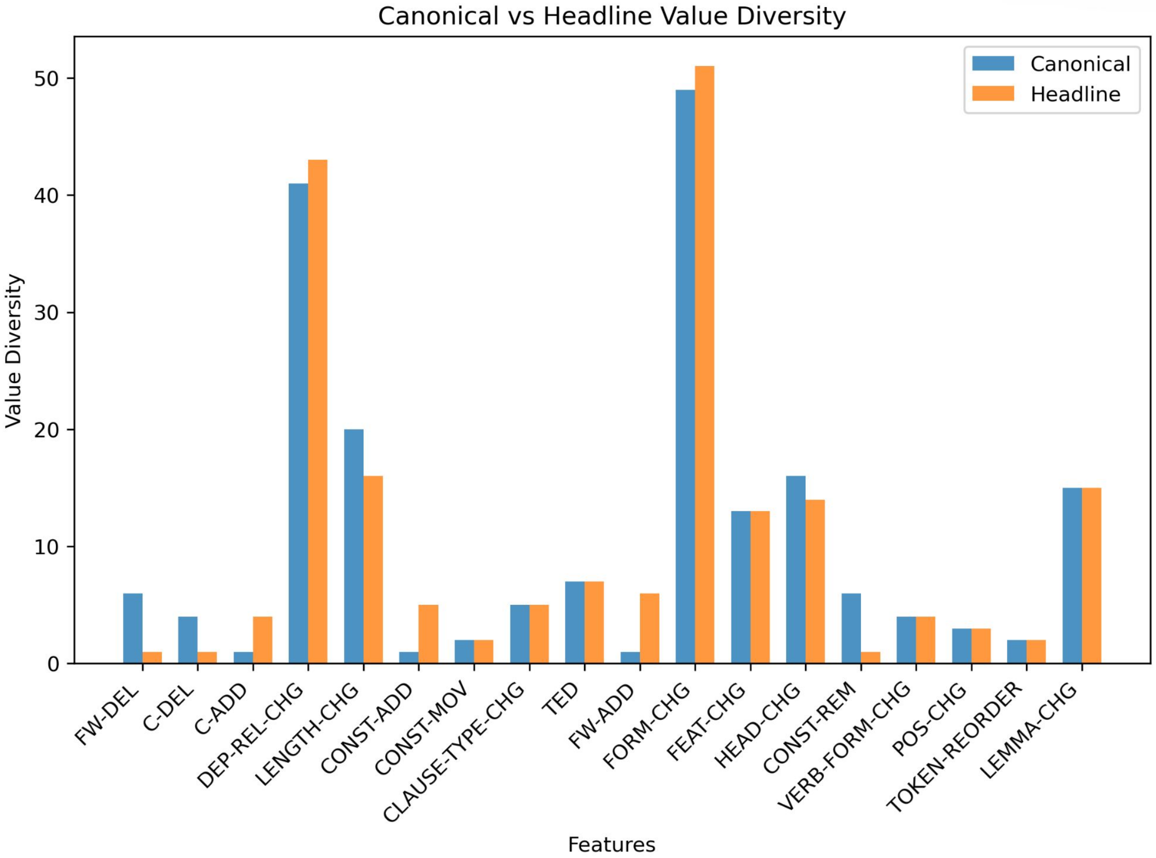


	C-ADD (Feature)	C-DEL (Feature)	CLAUSE-TYPE-CHG (Feature)	CONST-ADD (Feature)	CONST-MOV (Feature)	CONST-REM (Feature)	DEP-REL-CHG (Feature)	FEAT-CHG (Feature)	FORM-CHG (Feature)	FW-ADD (Feature)	FW-DEL (Feature)	HEAD-CHG (Feature)	LEMMA-CHG (Feature)	LENGTH-CHG (Feature)	POS-CHG (Feature)	TED (Feature)	TOKEN-REORDER (Feature)	VERB-FORM-CHG (Feature)
Hindustan-Times	690	1405	3408	114	13099	476	11759	108	61	266	3210	280	17	1137	84	1142	3	13
dependency	690	1405	0	0	0	0	11759	108	61	266	3210	280	17	1137	84	0	3	13
constituency	0	0	3408	114	13099	476	0	0	0	0	0	0	0	0	0	1142	0	0

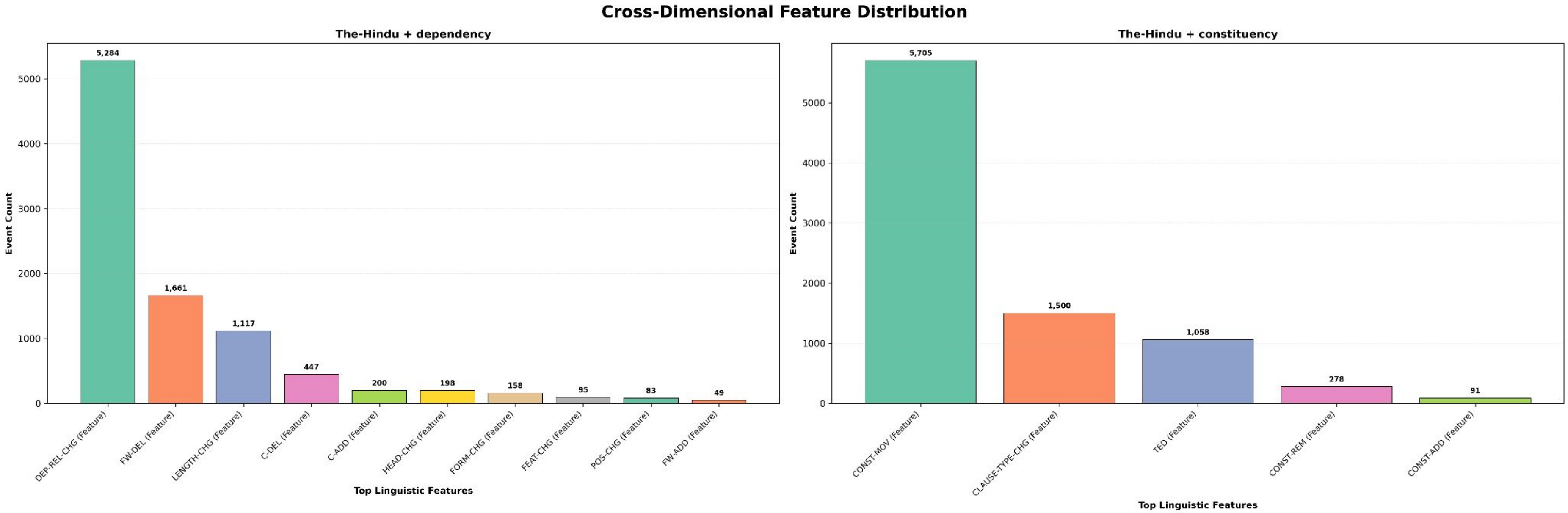
Global F-V Distribution: HT



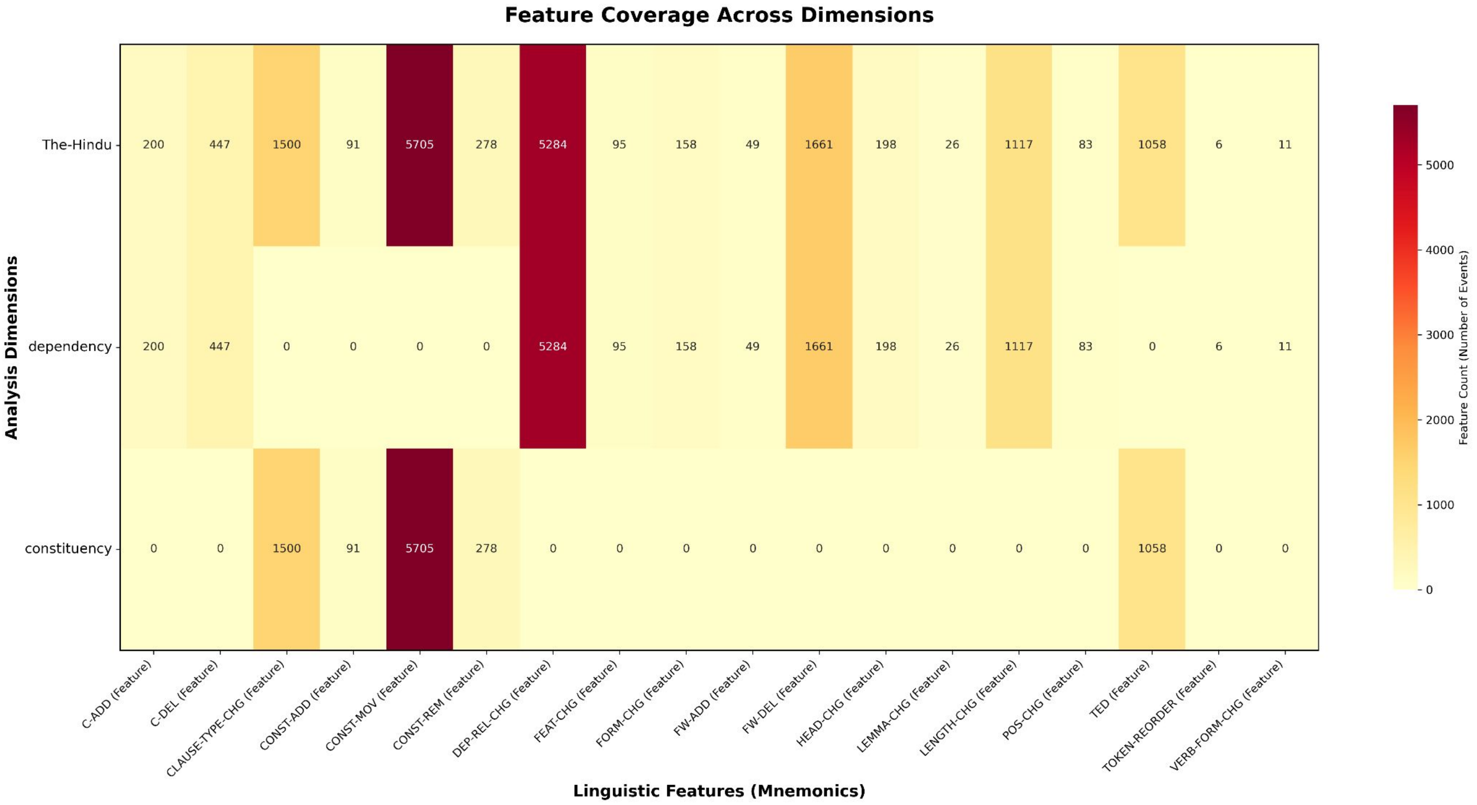
Canonical-Reduced Value Diversitv: HT



Cross-Dimensional Feature-Value Distribution: TH

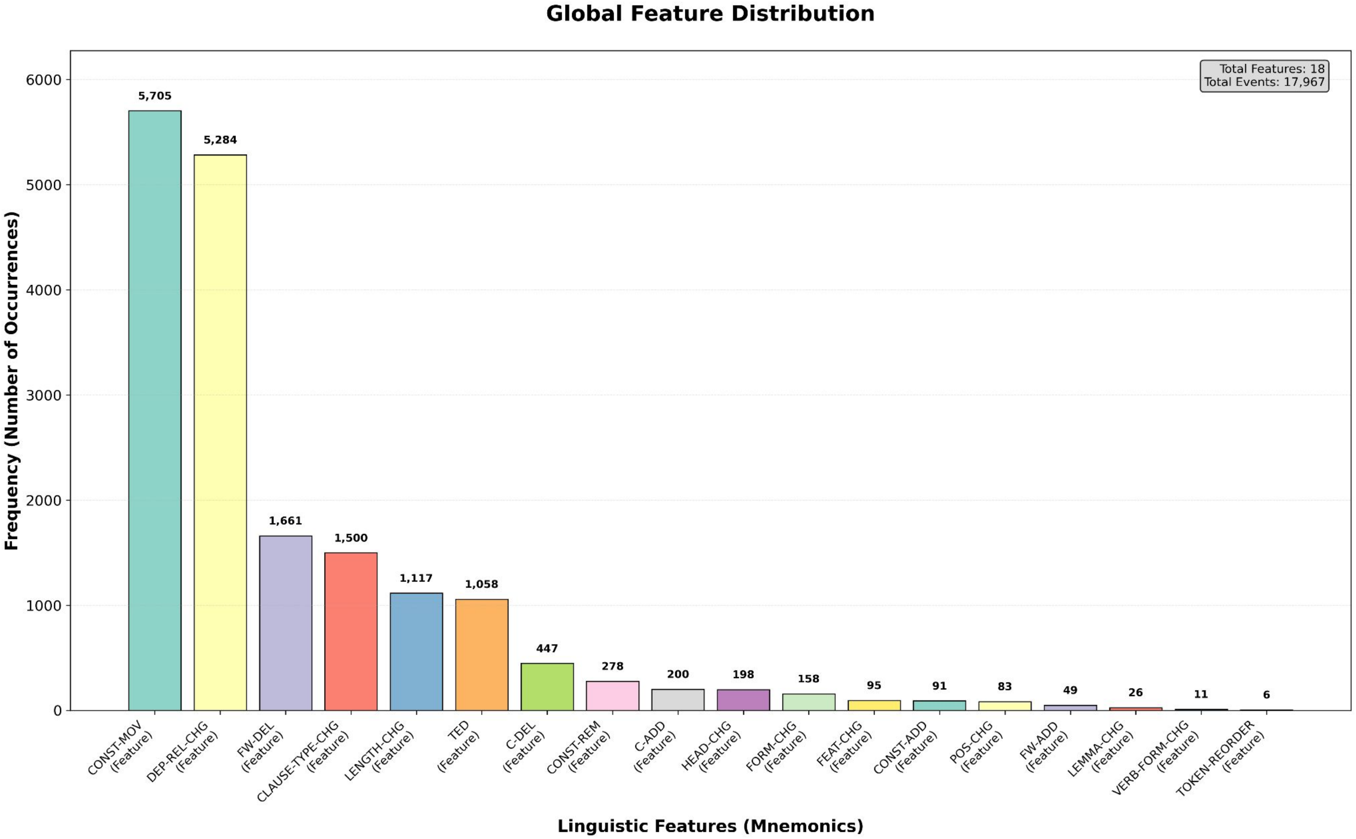


Feature-Value Pair Coverage Across Dimensions: TH

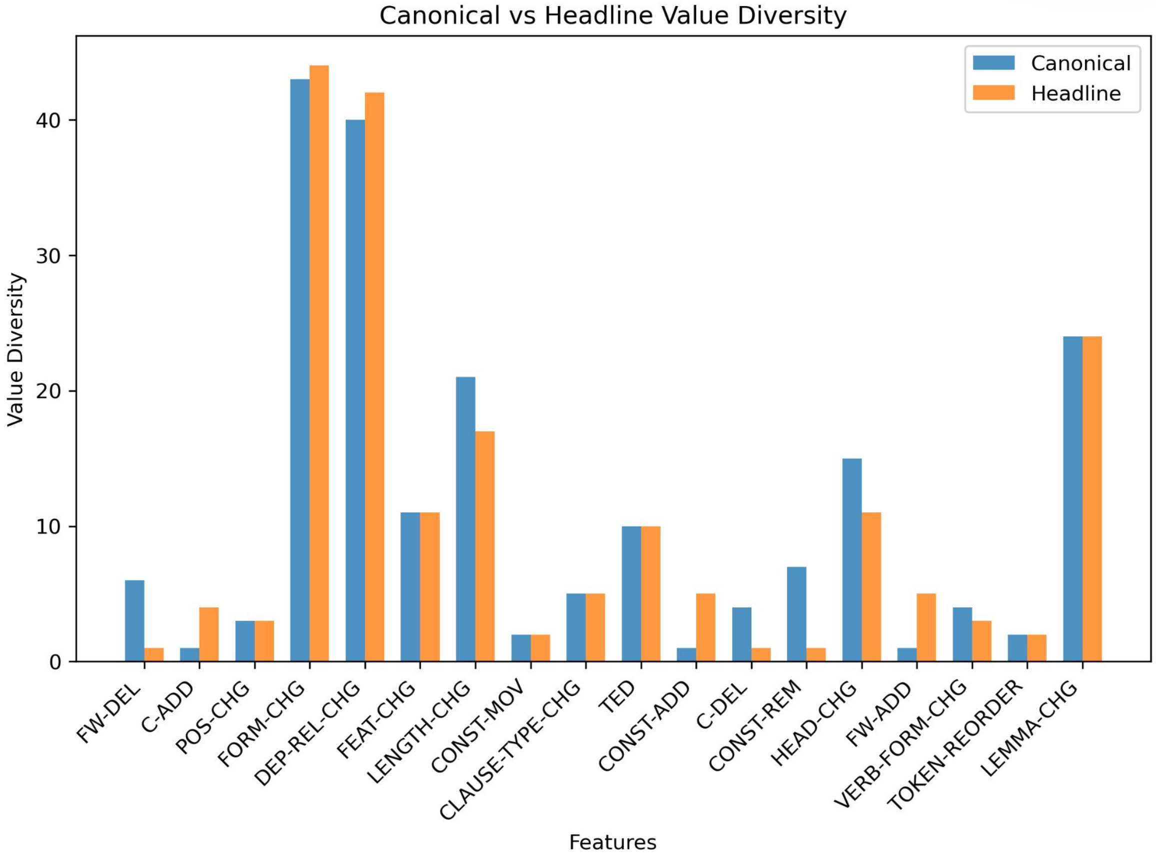


Analysis Dimensions	C-ADD (Feature)	C-DEL (Feature)	CLAUSE-TYPE-CHG (Feature)	CONST-ADD (Feature)	CONST-MOV (Feature)	CONST-REM (Feature)	DEP-REL-CHG (Feature)	FEAT-CHG (Feature)	FORM-CHG (Feature)	FW-ADD (Feature)	FW-DEL (Feature)	HEAD-CHG (Feature)	LEMMA-CHG (Feature)	LENGTH-CHG (Feature)	POS-CHG (Feature)	TED (Feature)	TOKEN-REORDER (Feature)	VERB-FORM-CHG (Feature)
The-Hindu	200	447	1500	91	5705	278	5284	95	158	49	1661	198	26	1117	83	1058	6	11
dependency	200	447	0	0	0	0	5284	95	158	49	1661	198	26	1117	83	0	6	11
constituency	0	0	1500	91	5705	278	0	0	0	0	0	0	0	0	0	1058	0	0

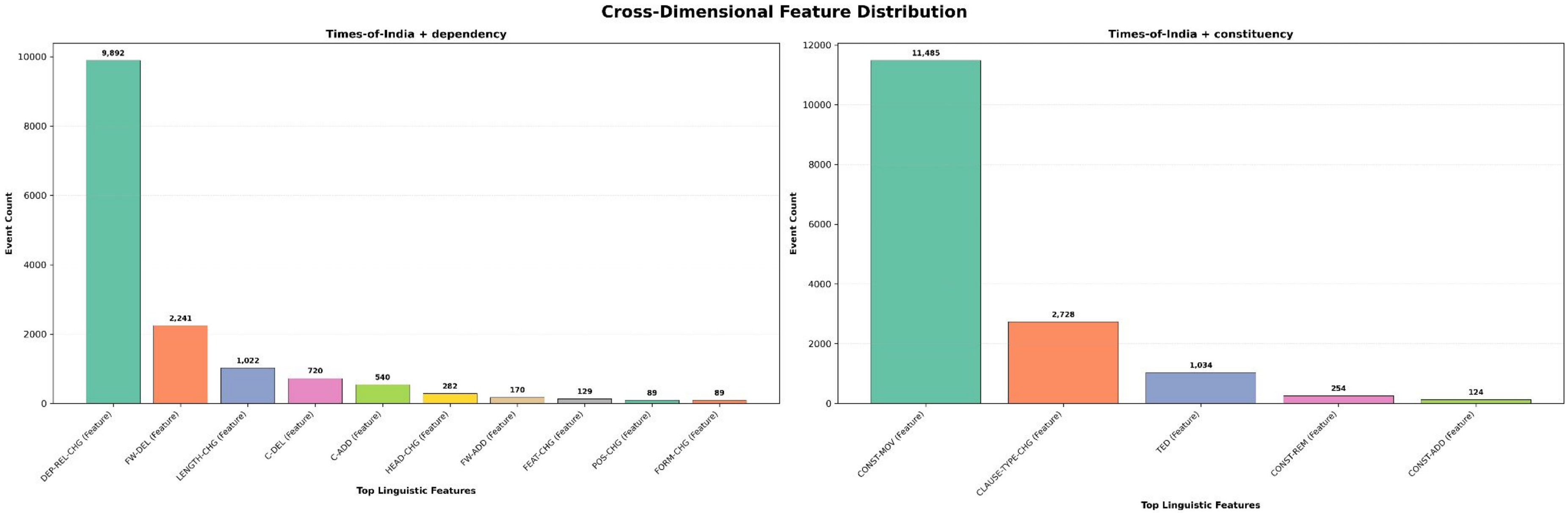
Global F-V Distribution: TH



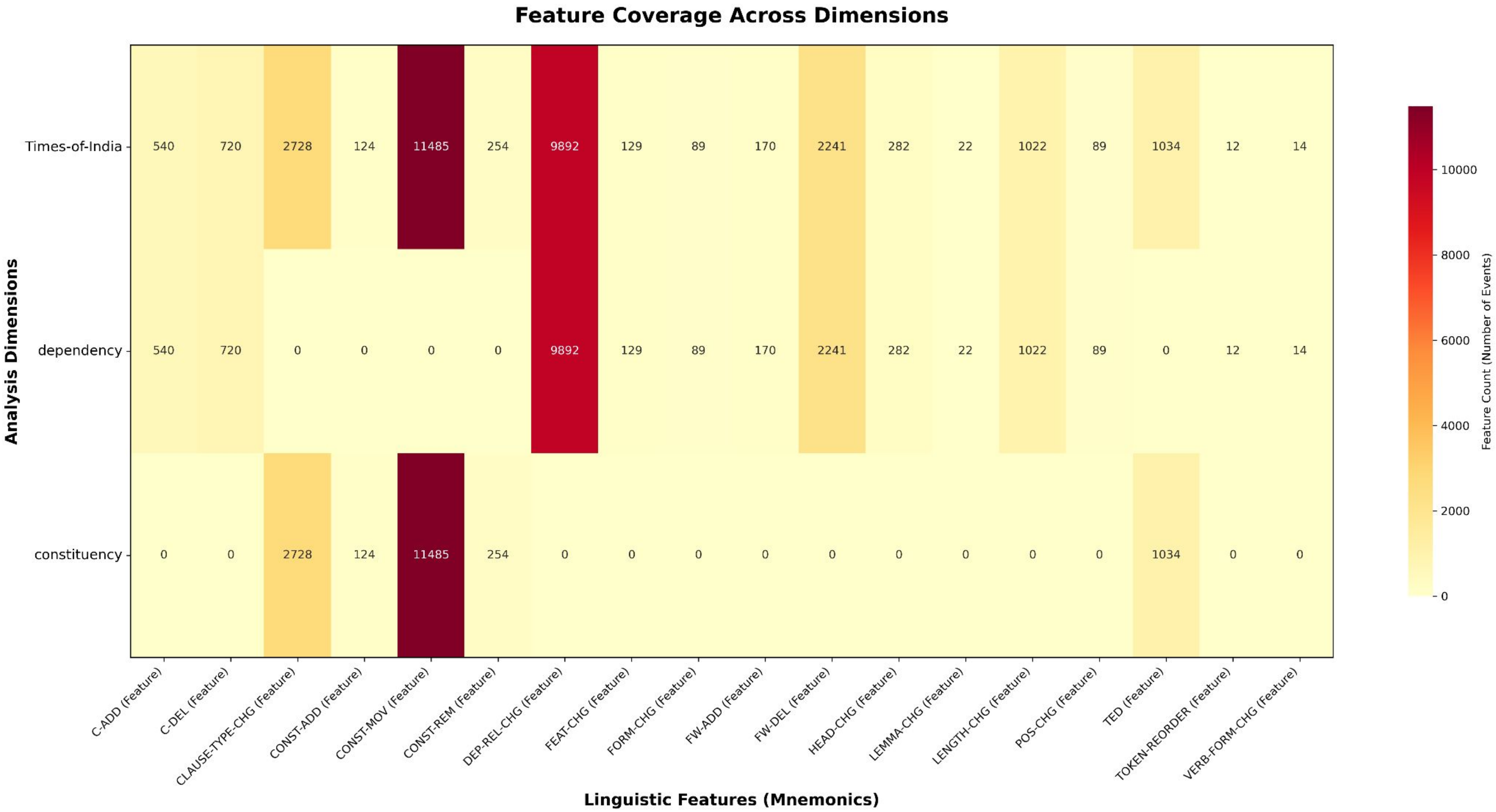
Canonical-Reduced Value Diversity: TH



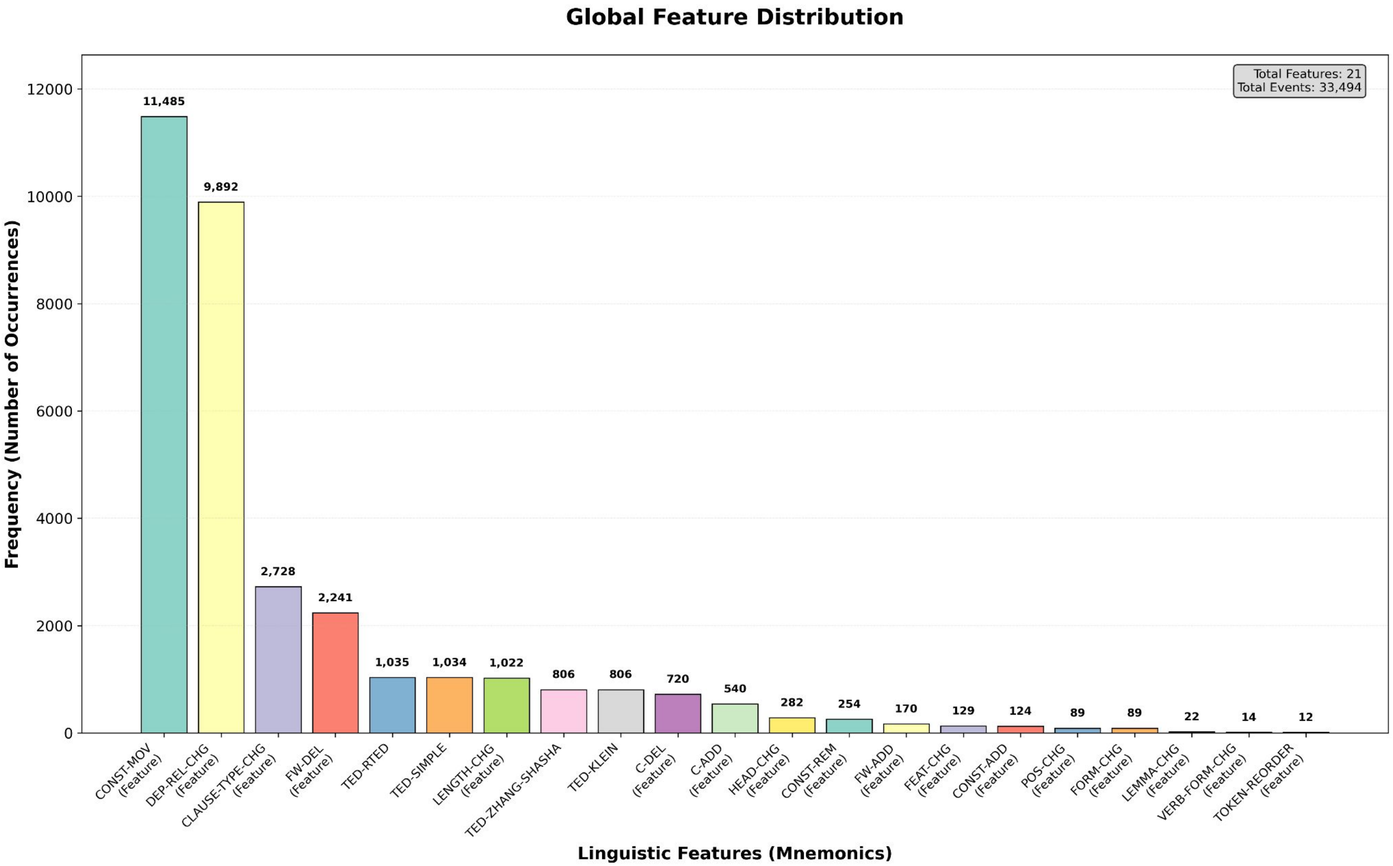
Cross-Dimensional Feature-Value Distribution: TOI



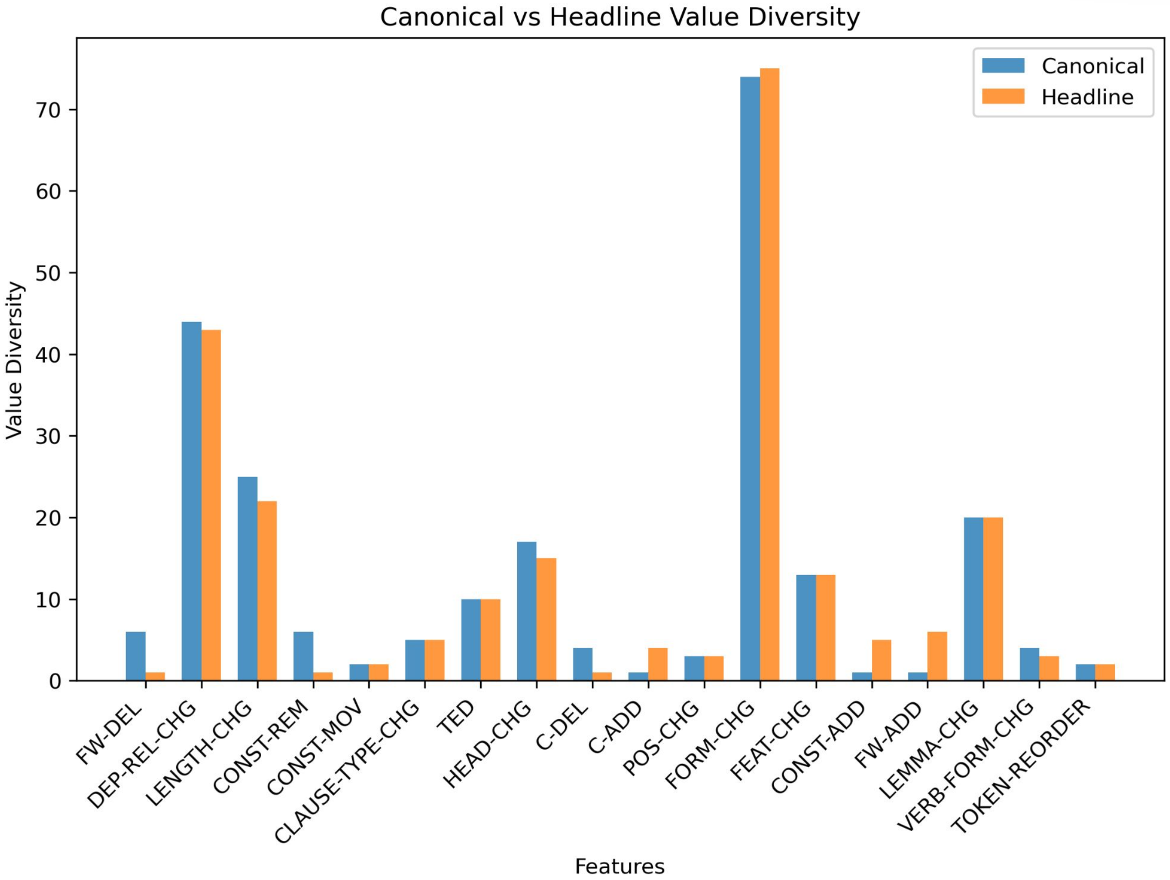
Feature-Value Pair Coverage Across Dimensions: TOI



Top Global F-V Distribution: TOI



Canonical-Reduced Value Diversity: TOI



Observations from the Results: Refined Schema

- As can be seen from the previous visualizations
  - Global (all newspaper data combined)
  - And individual newspapers
- Results are almost the same in terms of visualizations
- This means the register differences are
  - Systematic, irrespective of the newspaper
  - The same kinds of F-V differences appear
  - And in the same distributions

Conclusion

- The results validate the past theoretical studies
  - About register differences
- Same kinds of differences observed from quantitative comparison
  - And analysis
- Visualizations demonstrate this very clearly
- The experimental design can be used for other registers
- Results likely to be generalizable
  - But need more detailed checking in future work

Thank You!