Principal Parts Detection for Computational Morphology: Task, Models and Benchmark

Anonymous ACL submission

Abstract

Principal parts of an inflectional paradigm, defined as the minimal set of paradigm cells required to deduce all others, constitute an important concept in theoretical morphology. This concept, which outlines the minimal memorization needed for a perfect inflector, has been largely overlooked in computational morphology despite impressive advances in the field over the last decade. In this work, we posit PRINCIPAL PARTS DETECTION as a computational task and construct a multilingual dataset of verbal principal parts covering ten languages, based on Wiktionary entries. We evaluate an array of PRINCIPAL PARTS DETECTION methods, all of which follow the same schema: characterize the relationships between each pair of inflectional categories, cluster the resulting vector representations, and select a representative of each cluster as a predicted principal part. Our best-performing model, based on Edit Script between inflections and using Hierarchical K-Means, achieves an F1 score of 55.05%, significantly outperforming a Random Baseline of 21.20%. While our results demonstrate that some success is achievable, further work is needed to thoroughly solve PRINCIPAL PARTS DETECTION, a task that may be used to further optimize inputs for morphological inflection, and to promote research into the theoretical and practical importance of a compact representation of morphological paradigms.

1 Introduction

Morphological analysis is essential for understanding natural language, particularly in languages with complex inflectional systems. In both linguistic theory and language pedagogy, the concept of *principal parts* plays a central role in structuring and simplifying inflectional paradigms (Finkel and Stump, 2007; Stump and Finkel, 2013). Principal parts form the minimal subset of paradigm cells from which all other forms can be systematically derived.

By identifying these key forms, principal parts provide a compact representation of inflection tables and facilitate the analysis of morphologically rich languages. Despite their theoretical significance, the detection of principal parts remains largely unexplored in computational morphology. While they have inspired research in inflection and reinflection (Cotterell et al., 2017; Liu and Hulden, 2020), they are rarely used explicitly. Most computational approaches instead rely on a single citation form, the lemma (Cotterell et al., 2016; Goldman et al., 2023), or select input forms randomly (Cotterell et al., 2016; Kann et al., 2017). This reliance on suboptimal input representations overlooks the potential of principal parts as a more efficient foundation for inflectional modeling.

In this work, we introduce PRINCIPAL PARTS DETECTION as a formal task within computational morphology. Given a large collection of inflection tables, the goal is to determine which paradigm cells constitute the minimal principal-part set. Crucially, inflection tables typically contain standard morphological annotations but are not explicitly labeled with principal parts, making this an unsupervised learning problem. To promote research in this area, we deliver a standardized dataset covering the verbal paradigms of ten diverse languages. We sourced Principal parts for each language from online dictionaries, where they are often listed to aid language learners, and obtained full inflection tables from UniMorph (Batsuren et al., 2022).

We develop several computational approaches for PRINCIPAL PARTS DETECTION, leveraging the defining property of principal parts: their predictable and systematic relationships with other forms in the paradigm. Our models characterize inter-cell similarity and cluster inflected forms into *sub-paradigms*, selecting a representative cell from each sub-paradigm as candidate principal parts. We explore different methods for *characterizing* inter-cell relations, including Edit Distance, Edit

Script, and Reinflection Accuracy, and we experiment with *clustering* techniques such as Affinity Propagation and a modified K-Means algorithm. Our best-performing system, using Edit Script similarity measure + Hierarchical K-Means clustering, achieves an average F1 score of 55.05% across the ten languages in our dataset, significantly outperforming a Random Baseline of 21.20%.

By formalizing PRINCIPAL PARTS DETECTION as a computational task, we lay the groundwork for future research on more efficient morphological representations. To the best of our knowledge, this is the first work to deliver a standardized benchmark of PRINCIPAL PARTS DETECTION alongside a fully-operational detection framework. Successfully solving this task could enhance applications in morphological inflection and analysis by providing more informative input forms. Our findings suggest that principal parts can be computationally identified with reasonable accuracy, but further improvements are necessary to fully realize their potential.

2 The PRINCIPAL PARTS DETECTION Task and Dataset

The PRINCIPAL PARTS DETECTION Task. The task of PRINCIPAL PARTS DETECTION is defined as identifying the minimal set of cells within a paradigm that, when known, allow the derivation of all other paradigm forms. For instance, in English, the principal parts of the verbal paradigm are the cells corresponding to the infinitive, simple past and past participle (for example, *eat*, *ate*, and *eaten*), as these forms are not predictable from one another, especially for strong verbs. On the other hand, the forms corresponding to the present participle and the 3rd person singular present are deterministically predictable from the infinitive and they therefore provide no additional information for inflection if the infinitive is known.

Formally, the task of PRINCIPAL PARTS DETECTION is defined given a language L, a paradigm $P \in L$, and a large set of inflection tables $T = \left\{t^L_{P,1}, t^L_{P,2}, \ldots, t^L_{P,n}\right\}$ that belong to that paradigm. The goal is to identify a minimal set of cells $C_{PP} \subseteq P$ from which all other cells in all inflection tables of P can be accurately deduced.

The PRINCIPAL PARTS DETECTION Dataset. In order to empirically assess methods for the detection of principal parts, we first need to have a dataset to evaluate against. To this end, we constructed the PRINCIPAL PARTS DETECTION dataset, containing ten typologically diverse languages, where every paradigm in every language is characterized by a set of target principal parts that systems can be evaluated against.

The input side of the task contains complete inflection tables, based on the UniMorph corpus (Batsuren et al., 2022), which provides inflection tables for 168 languages organized by lexeme and morpho-syntactic features. We sourced gold principal parts — that are the desired output — from a combination of Wiktionary and other trusted online dictionaries or language teaching websites. Based on the availability of data sources for both input and output, we selected ten typologically diverse languages: Hebrew, English, French, German, Spanish, Danish, Swedish, Finnish, Turkish and Latin.

The dataset preparation process involved normalizing the data for consistency across languages. Redundant and derivational forms were excluded, leaving only core inflectional forms. Inconsistent feature sets were removed, and problematic entries from the original sources were manually corrected to ensure a reliable dataset (for more details, see Appendix A). The PRINCIPAL PARTS DETECTION dataset provides a strong foundation for computational models, bridging linguistic theory and practical applications. By curating this multilingual dataset, we ensure a robust resource for future research in morphological inflection. The next section shifts focus to computational methods for detecting principal parts, drawing on the linguistic insights outlined in the literature.[1]

3 Translating Linguistic Insights into Computational Methods

The core linguistic principle underlying PRINCIPAL PARTS DETECTION is that principal parts encapsulate the implicative relationships that exist between cells in inflectional paradigms, allowing a small set of cells to reconstruct the full inflectional table. To translate this principle into a computationally tractable problem, we frame PRINCIPAL PARTS DETECTION as the automatic identification of a minimal, generative subset of paradigm cells that can generate all other cells via these implicative relationships.

We hypothesize that cells of different feature sets in an inflection table exhibit measurable simi-

[1]The data is publicly available in https://www.will.be.released/upon.acceptance.

larities in their realized surface forms, that in turn reflect morphological and structural relationships between these feature-set cells. By capturing patterns of inter-dependence between cells, we approximate the implicative structure of a paradigm *without* relying on explicit linguistic annotations of principal parts. To systematically model these inter-dependencies, we introduce the notion of a *sub-paradigm*, which is essentially a sub-set of paradigm cells with shared implicative properties, and define distinct areas-of-interdependence within a paradigm. Although sub-paradigms are not a formal linguistic concept, they provide a structured way to model implicative relationships computationally, facilitating the detection of principal parts.

This conceptualization leads to a three-phased methodology for PRINCIPAL PARTS DETECTION. First, we *characterize* the relationships between pairs of cells by computing similarity measures that capture their surface and structural dependencies. Next, we *cluster* inter-related cells into sub-paradigms, each of which will be represented by a single principal part in the final set. Finally, we *select* one representative feature set per sub-paradigm as its designated principal part, ensuring maximal coverage of the paradigm cells with minimal redundancy. The instantiation of these *(i) characterization, (ii) clustering* and *(iii) candidate selection* phases gives rise to a host of PRINCIPAL PARTS DETECTION concrete implementations that we can define and empirically assess — as we discuss next.

4 Framework and Task Empirical Design

The PRINCIPAL PARTS DETECTION framework we propose here is composed of three interconnected stages: characterization, clustering, and principal parts selection, each implemented using well-defined computational methods. These stages operate independently, meaning that different configurations of the framework can mix and match methods in seeking the best combination. Let us briefly review the computational models we consider for the different phases.

4.1 Characterization: Quantifying Relationships Between Feature Sets

The characterization stage quantifies the relationships between paradigm cells by computing numerical similarity scores between them. This work explores three distinct characterization methods, offering a different perspectives on the relation between cells.

Edit Distance A metric that measures surface-level similarity between forms based on minimal edit operations — insertions, deletions, or substitutions — required to transform one form into another (Levenshtein, 1966). This method is implemented by computing the average Edit Distance from each feature set to all others (calculated across all their surface realizations), treating one as the source and the rest as destinations. The resulting vector representations store these averaged distances, capturing the surface-level similarity between feature sets. Pairs of paradigm cells with low Edit Distance scores exhibit orthographic overlap.

Edit Script A metric that captures transformational diversity by analyzing character-level transformations between paradigm cells. Unlike traditional Edit Script approaches (Wagner and Fischer, 1974; Myers, 1986), which focus on the exact sequence of operations needed to transform one string into another, this approach computes the number of unique character-level transformations observed across all surface realizations of each feature set pair. Each transformation is counted only once per feature set pair (calculated across all their surface realizations), capturing distinct transformational patterns rather than repeatedly occurring character changes. The result is a vector representation for each feature set pair, where each entry encodes the number of unique transformations required to convert one feature set to another, representing their transformational distance. This method provides insight into the variation in morphological transformations within a paradigm. Feature sets with lower transformation diversity may exhibit more stable morphological patterns, making them stronger principal part candidates. In contrast, higher transformation diversity may signal greater variability in inflectional behavior, which can affect predictability within the paradigm.

Reinflection Accuracy A metric that evaluates the functional predictability of feature sets. It leverages a neural reinflection model trained to generate a target form given a source form and the morpho-syntactic features of the target. Unlike edit-based methods that focus on surface similarity and transformational diversity, Reinflection Accuracy captures the functional dependencies between feature sets, reflecting their predictive capacity within a paradigm.

Reinflection accuracy is particularly effective in languages with complex inflectional systems, where orthographic similarity alone is not a reliable predictor of implicative relationships. By capturing functional dependencies rather than surface transformations, it provides a direct measure of a feature set's ability to generate other forms. However, its performance depends on training data quality and resource availability. In low-resource settings, data sparsity may lead to biased results, and the approach is computationally intensive, as it requires training multiple models—one model per feature set. Despite these challenges, its ability to model functional predictability makes it a valuable tool for identifying feature sets that serve as principal parts, particularly in morphologically complex languages.

Each characterization method produces a similarity table, where rows represent source feature sets and columns represent target feature sets, encoding pairwise relationships (see Appendix B). Before clustering, all similarity matrices are standardized by removing the mean and scaling to unit variance to ensure comparability across methods. These standardized characterization tables form the foundation for the clustering stage.

4.2 Clustering: Structuring Feature Sets into Sub-Paradigms

The clustering stage groups feature sets based on their quantified relationships, approximating sub-paradigms that reflect the internal organization of inflectional paradigms. The framework implements two clustering algorithms, each offering different advantages. As with characterization, only one clustering algorithm is used at a time.

Affinity Propagation A message-passing clustering algorithm that dynamically determines the number of clusters based on pairwise similarity scores (Frey and Dueck, 2007). Unlike traditional clustering methods, it does not require a predefined number of clusters. Instead, it iteratively updates responsibility and availability values, which determine how well a feature set serves as an exemplar (cluster center), until the algorithm converges on a final set of exemplars. This property makes it particularly well-suited for paradigms with high morphological variability. The algorithm is implemented using scikit-learn's AffinityPropagation module, with similarity scores computed as negative squared Euclidean distances. The preference parameter is set to the median similarity value, allowing clusters to emerge naturally. Additional parameters include a convergence iteration limit of 30 and a random state value of 10.

Hierarchical K-Means A hierarchical variant of K-Means that recursively partitions feature sets into two clusters per iteration until a well-defined clustering structure is reached. The stopping criterion is determined using the Calinski–Harabasz Index (CHI) (Caliński and Harabasz, 1974), which evaluates clustering quality by comparing between-cluster dispersion to within-cluster cohesion. At each step, the CHI is computed across the entire clustering structure to evaluate how well-separated the clusters are relative to their internal cohesion. To prevent over-segmentation, clustering stops if the number of clusters in the new best CHI solution exceeds that of the previous best CHI solution by more than one cluster. The algorithm is implemented using scikit-learn's KMeans module with a random state value of 10.

By grouping feature sets into paradigm subsets, the clustering stage provides a data-driven approximation of sub-paradigms. The resulting clusters serve as inputs for the principal parts selection stage.

4.3 Principal Parts Selection: Identifying Representative Feature Sets

The principal parts selection stage finalizes the PRINCIPAL PARTS DETECTION framework by transforming clusters into a compact and generative summary of the paradigm. This stage selects one representative feature set per cluster, encapsulating its defining structural and transformational relationships. These feature sets collectively constitute the principal parts, providing comprehensive coverage while maintaining a balance between compactness and predictive capacity.

Concretely, we use the Minimum Average Inflectional Length criterion. That is, the feature set with the minimal average inflectional length in its cluster, calculated across all its surface realizations, is chosen as the principal part. This ensures that the selected feature set is both efficient and central within its cluster. This selection criterion aligns with a linguistic insight that shorter inflectional paths often correspond to forms that are central within the paradigm, making them structurally significant within inflectional systems.

5 Experimental Setup and Results

We conduct a series of experiments to evaluate the effectiveness of the PRINCIPAL PARTS DETECTION framework across ten typologically diverse languages. The evaluation compares six model configurations, each formed by pairing one of three characterization methods—Edit Distance, Edit Script, and Reinflection Accuracy—with one of two clustering algorithms—Affinity Propagation and Hierarchical K-Means. To establish a performance threshold, we include a Random Baseline, which selects principal parts at random.

5.1 Dataset

The experiments are conducted on the PRINCIPAL PARTS DETECTION dataset, which comprises ten typologically diverse languages, divided into a development set (Hebrew, English, French, German, and Spanish) and a test set (Danish, Swedish, Finnish, Turkish, and Latin).

The development set represents varied morphological structures. Hebrew exhibits synthetic morphology, encoding multiple grammatical elements within single word forms. English, in contrast, is analytic, primarily relying on word order and auxiliary constructions for grammatical relationships. French and Spanish, as fusional languages, encode tense, mood, and person within single inflections, albeit with varying degrees of regularity. German, a hybrid case, incorporates both fusional and analytic morphological characteristics, presenting distinct patterns for analysis. This linguistic diversity ensures that models are trained on paradigms with different degrees of morphological richness, regularity, and complexity.

The test set is designed to assess generalization across languages with distinct inflectional systems. Finnish and Turkish exemplify agglutinative morphology, where grammatical meaning is expressed through concatenative morphemes. Latin, a highly inflected classical language, provides a rigorous test case for evaluating the models' ability to handle case, number, and gender distinctions. Danish and Swedish, characterized by relatively regular inflectional systems, contribute typological variety while testing the models' robustness in less complex paradigms.

By structuring the dataset to reflect a wide range of linguistic variation, this division ensures a comprehensive evaluation of the framework's adaptability to diverse morphological systems and its ability to generalize across typologically distinct languages.

5.2 Evaluation Metric

To evaluate model effectiveness, we use the F1 score, which balances precision and recall to assess both accuracy and completeness in PRINCIPAL PARTS DETECTION.

In addition to reporting F1 scores, we compare model performance against a Random Baseline, which selects principal parts randomly within each paradigm. Given a paradigm with x feature sets and y gold principal parts, the probability of randomly selecting a correct principal part is $\frac{y}{x}$. Since the baseline selects y principal parts, the expected number of correct predictions is $y \times \frac{y}{x} = \frac{y^2}{x}$. From this, the expected precision, recall, and F1 score are all: $F1 = \frac{y}{x}$. Since principal parts are inherently sparse within most paradigms, the Random Baseline represents a challenging threshold. Models that significantly exceed this score demonstrate an ability to detect principal parts systematically rather than relying on chance.

5.3 Reinflection Settings

For models utilizing Reinflection Accuracy, we train a separate reinflection model for each feature set, treating it as the source while all other feature sets serve as targets. The model is based on the Base LSTM architecture (Goldman et al., 2021), a character-based sequence-to-sequence model comprising a one-layer bidirectional LSTM encoder and a one-layer unidirectional LSTM decoder with a global soft attention layer (Bahdanau et al., 2014). Each model is trained for 50 epochs, optimizing categorical cross-entropy.

The dataset is split 70%-30%, ensuring that test lexemes are unseen during training. Each feature set is trained using a dedicated dataset, where it serves as the source inflection across different lexemes. Since each feature set is evaluated on its ability to generate all other feature sets within the paradigm, corresponding test sets are created—one per target feature set.

Each trained model is evaluated on how accurately it inflects from its assigned source feature set to each target feature set. The resulting accuracy scores form representation vector, capturing a feature set's proficiency in generating others. Feature sets with high Reinflection Accuracy scores demonstrate strong predictive capacity, making them effective candidates for principal parts.

Model	Algorithmic Evaluation
Random Baseline	21.20
Edit Distance + Affinity Propagation	31.29
Edit Distance + Hierarchical K-Means	32.51
Reinflection Accuracy + Hierarchical K-Means	42.43
Edit Script + Affinity Propagation	44.62
Reinflection Accuracy + Affinity Propagation	45.56
Edit Script + Hierarchical K-Means	**55.05**

Table 1: Averaged F1 scores of PRINCIPAL PARTS DETECTION models across ten languages. The table compares different model configurations, highlighting the best-performing model.

Model	Hebrew	English	French	German	Spanish	Danish	Swedish	Finnish	Turkish	Latin
Random Baseline	20.68	60.00	14.28	16.66	2.53	62.50	26.30	2.48	28.00	6.25
Edit Distance + Affinity Propagation	33.30	66.70	37.50	46.20	15.40	57.10	40.00	0.00	0.00	16.70
Edit Distance + Hierarchical K-Means	25.00	57.10	44.40	44.40	0.00	57.10	57.10	0.00	0.00	40.00
Reinflection Accuracy + Hierarchical K-Means	25.00	85.70	44.40	28.60	50.00	57.10	43.50	50.00	0.00	40.00
Edit Script + Affinity Propagation	50.00	80.00	54.50	66.70	36.40	50.00	60.00	23.50	6.90	18.20
Reinflection Accuracy + Affinity Propagation	36.40	80.00	26.70	60.00	16.70	75.00	75.00	46.20	17.40	22.20
Edit Script + Hierarchical K-Means	**50.00**	80.00	**54.50**	60.00	**50.00**	72.70	60.00	33.30	**50.00**	**40.00**

Table 2: Language-specific F1 scores of PRINCIPAL PARTS DETECTION models across ten languages. The table highlights variations in model effectiveness across different morphological typologies.

5.4 Results

Table 1 presents the average F1 scores across ten languages, providing a comparative evaluation of model performance. All models outperform the Random Baseline, which achieves the lowest F1 score of 21.20%. The best-performing model, Edit Script + Hierarchical K-Means, achieves an F1 score of 55.05%, demonstrating its ability to effectively capture and cluster morphological patterns across diverse languages.

Reinflection Accuracy models perform competitively, with F1 scores of 45.56% (Affinity Propagation) and 42.43% (Hierarchical K-Means). In contrast, Edit Distance-based models yield lower scores of 31.29% and 32.51%, indicating that surface-level similarity alone is insufficient for PRINCIPAL PARTS DETECTION.

Overall, all tested methods surpass the Random Baseline by at least 10.09 points, with the best-performing model exceeding it by 33.85 points. These results confirm the effectiveness of the proposed methodology, demonstrating a substantial improvement over random selection.

Table 2 provides a language-specific breakdown of F1 scores, offering further insight into model performance across different morphological typologies. Edit Script + Hierarchical K-Means achieves top performance in Hebrew, French, Spanish, Turkish, and Latin, confirming its adaptability across different morphological systems. Reinflection Accuracy-based models perform particularly well in English, Spanish, Finnish, and Swedish, suggesting that functional predictability is well-suited for these languages.

Interestingly, while Reinflection Accuracy + Affinity Propagation ranks second overall (45.56%), it does not consistently outperform other models across all languages. In Danish and Swedish, its relatively strong results suggest an advantage in regular inflectional systems where paradigmatic structures are highly predictable. Conversely, in fusional languages like Spanish, where single inflections encode multiple grammatical features, it faces challenges in PRINCIPAL PARTS DETECTION.

In contrast, Edit Distance-based models fail to rank highest in any language, reinforcing the conclusion that surface-level similarity alone is insufficient for PRINCIPAL PARTS DETECTION. These findings emphasize the importance of selecting appropriate characterization methods based on linguistic properties and show that transformational diversity (Edit Script) and functional predictability (Reinflection Accuracy) are particularly effective strategies.

6 Analysis

We analyze how methodological factors shape model performance, focusing on transformations in characterization data and the effectiveness of clustering strategies. This evaluation highlights structural patterns influencing clustering quality and examines the extent to which clustering results align with ideal principal parts selection.

6.1 Transpose Ablation: Evaluating the Impact of Data Orientation

The Transpose Ablation study investigates whether swapping the rows and columns of the characterization tables influences clustering quality and principal parts selection. This transformation is particularly relevant for Reinflection Accuracy, where the original tables encode directional relationships—rows indicate how easily a feature set can inflect from itself to others, while columns represent the reverse relationship. By transposing these tables, we test whether an alternative structural alignment improves performance.

Model	Transpose	Algorithmic Evaluation
Reinflection Accuracy + Affinity Propagation	✗	**45.56**
	✓	44.05
Reinflection Accuracy + Hierarchical K-Means	✗	42.43
	✓	**43.14**

Table 3: Algorithmic evaluation of Reinflection Accuracy models with and without transposition across ten languages. The table presents the averaged F1 scores for models before and after transposition, highlighting its varying impact depending on the clustering algorithm.

Model	Evaluation	
	Oracle	Algorithmic
Edit Distance + Affinity Propagation	40.08	31.29
Edit Distance + Hierarchical K-Means	50.57	32.51
Reinflection Accuracy + Affinity Propagation	58.78	45.56
Reinflection Accuracy + Hierarchical K-Means	65.64	42.43
Edit Script + Affinity Propagation	54.16	44.62
Edit Script + Hierarchical K-Means	**76.21**	**55.05**

Table 4: Oracle and Algorithmic evaluations of PRINCIPAL PARTS DETECTION models across languages. Oracle evaluation assumes perfect knowledge of principal parts, establishing an upper bound on performance, while Algorithmic evaluation reflects actual model performance.

Model	Transpose	Evaluation	
		Oracle	Algorithmic
Reinflection Accuracy + Affinity Propagation	✗	**58.78**	**45.56**
	✓	58.51	44.05
Reinflection Accuracy + Hierarchical K-Means	✗	65.64	42.43
	✓	**67.70**	**43.14**

Table 5: Oracle and Algorithmic evaluations of Reinflection Accuracy before and after transposition. The table examines how transposition affects clustering quality under both ideal (Oracle) and algorithmic conditions.

Transposition is applied only to Reinflection Accuracy models, as Edit Distance and Edit Script methods generate symmetric similarity matrices, making transposition redundant. We evaluate two models: Reinflection Accuracy + Affinity Propagation and Reinflection Accuracy + Hierarchical K-Means, comparing their performance before and after transposition.

The results in Table 3 show that transposition affects models differently. Reinflection Accuracy + Affinity Propagation experiences a slight decrease in performance (45.56% $\rightarrow$ 44.05%), while Reinflection Accuracy + Hierarchical K-Means improves marginally (42.43% $\rightarrow$ 43.14%). This suggests that transposition does not universally enhance clustering effectiveness and that its impact depends on the underlying clustering strategy.

Despite the minor improvement in Hierarchical K-Means, transposed results are excluded from the main evaluation due to their limited effect and misalignment with the principal parts definition. Since original (non-transposed) feature sets encode generative properties crucial for inflection, preserving this structure remains preferable. These findings suggest that alternative data transformations, better aligned with the linguistic task, may offer greater benefits.

6.2 Oracle Evaluation

To assess the theoretical upper limit of model performance, we conduct an Oracle evaluation, where principal parts are selected with perfect knowledge rather than through clustering. This evaluation distinguishes clustering effectiveness from principal parts selection quality, highlighting areas for improvement.

Table 4 reveals substantial gaps between Oracle and Algorithmic scores, underscoring clustering limitations and principal parts selection inefficiencies. Edit Script + Hierarchical K-Means achieves the highest Oracle score (76.21%), confirming strong clustering performance. However, the 21.16-point gap suggests that principal parts selection remains a limiting factor.

Conversely, Edit Distance + Affinity Propagation exhibits the lowest Oracle score (40.08%), indicating fundamental challenges in clustering feature sets meaningfully. Reinflection Accuracy + Hierarchical K-Means shows a particularly large Oracle-Algorithmic gap (65.64% $\rightarrow$ 42.43%), highlighting that while clustering is effective, principal parts selection still requires refinement.

These findings emphasize the importance of optimizing both clustering effectiveness and principal parts selection to bridge the gap between Oracle and Algorithmic performance.

6.3 Interplay Between Transposition and Oracle Performance

Table 5 presents the impact of transposition on Reinflection Accuracy models under both Oracle and Algorithmic evaluations.

The results indicate that while transposition improves Oracle performance for Hierarchical K-Means (65.64% $\rightarrow$ 67.70%), it has a negligible effect on Algorithmic scores, indicating that while transposition enhances clustering under ideal conditions, it does not meaningfully improve principal parts selection. Additionally, Affinity Propagation exhibits sensitivity to data orientation, showing a slight decline in Oracle performance (58.78% $\rightarrow$ 58.51%), suggesting that its clustering mechanism relies on specific directional patterns that transposition may disrupt. Conversely, Hierarchical K-Means benefits from transposed data, likely due to its iterative refinement of clusters. However, since Algorithmic scores remain largely unchanged across models, these findings reinforce that refining selection heuristics, rather than adjusting data orientation, is the key to improving model performance.

7 Related Work

Early computational approaches to paradigm completion predominantly relied on the lemma as the central reference form, treating it as the sole input for generating full inflectional paradigms (Durrett and DeNero, 2013; Hulden, 2014; Nicolai et al., 2015; Ahlberg et al., 2015; Faruqui et al., 2016). However, Cotterell et al. (2017) highlighted the limitations of this approach, noting that forcing transformations to pass exclusively through the lemma can introduce unnecessary complexity. Instead, more flexible models leveraging multiple inflected forms have been proposed, allowing transformations to occur directly or via intermediary forms, rather than constraining them to a single privileged form. This shift aligns with the concept of principal parts, which constitute the minimal set of paradigm cells required to deduce all others (Finkel and Stump, 2007; Stump and Finkel, 2013).

Cotterell et al. (2017) introduced a directed graphical model that probabilistically generates missing inflected forms by modeling dependencies within paradigms. This approach enables the prediction of a form from multiple inflected forms rather than exclusively from the lemma. Around the same time, Kann et al. (2017) introduced multi-source reinflection, demonstrating that using multiple inflected forms as input improves accuracy. Their work explicitly references principal parts as a linguistic motivation, reinforcing the idea that certain forms within a paradigm hold stronger predictive capacity. Additionally, Cotterell et al. (2019) examined the structural complexity of inflectional paradigms, proposing a neural method for ordering paradigm slots based on their predictability—an indirect computational realization of the principal parts concept.

Liu and Hulden (2020) extended these ideas by reformulating morphological inflection as a Paradigm Cell Filling Problem (PCFP), where missing forms are inferred from a partially observed set of paradigm cells. While their work does not explicitly model principal parts, it aligns with their predictive role in improving inflectional accuracy, particularly in low-resource settings.

Despite these advancements, no prior work has proposed a systematic, data-driven approach to principal parts detection. Existing studies have either assumed pre-defined principal parts or incorporated them indirectly within broader inflectional tasks. In contrast, we introduce PRINCIPAL PARTS DETECTION as a formal computational task, developing a multilingual benchmark and a principled methodology for automatic PRINCIPAL PARTS DETECTION. By integrating linguistic insights with computational modeling, we establish a structured framework for principal parts detection.

8 Conclusions

This work introduces PRINCIPAL PARTS DETECTION as a computational task, formalizing the detection of principal parts within inflectional paradigms. We construct a multilingual dataset with ten typologically diverse languages, and develop a structured framework to automatically detect principal parts in their verbal diagrams.

our empirical evaluation shows that characterizing inter-cell relationships, clustering feature sets, and selecting representatives, offers a viable strategy for identifying principal parts. Our best-performing approach — Edit Script similarity with Hierarchical K-Means — achieves an F1 score of 55.05%, significantly surpassing the Random Baseline of 21.20%. However, results across models indicate that while clustering is effective in grouping related feature sets, principal parts selection remains a key bottleneck.

Beyond theoretical interest, solving PRINCIPAL PARTS DETECTION has practical implications for computational morphology. By identifying compact, generative subsets of paradigm forms, principal parts can be leveraged to optimize morphological inflection models, reduce annotation costs, and improve low-resource language modeling. The structured approach presented here lays the foundation for future advancements, underscoring the relevance of linguistic principles in shaping more efficient NLP methodologies.

Limitations

Despite the progress demonstrated in this study, several open challenges remain. Irregular paradigms, as seen in Latin, continue to pose difficulties, highlighting the need for methods that can better capture morphological unpredictability. Additionally, our reliance on UniMorph, while offering broad linguistic coverage, exposes inconsistencies that impact model generalization. More curated linguistic resources could improve dataset reliability and refine the evaluation of principal parts across languages.

Also, one could explore alternative clustering strategies that are better suited to morphological structures, such as graph-based methods or neural clustering approaches. Transformer-based models hold potential for capturing deeper morphological dependencies, offering an avenue for enhancing both clustering accuracy and principal parts selection. These challenges are beyond the scope of this paper and we reserve it to future work.

Our dataset currently includes only 10 languages. Expanding the dataset to include more morphologically rich and underrepresented languages, such as polysynthetic languages, would better capture typological diversity and will potentially further validate the robustness of PRINCIPAL PARTS DETECTION methods.

References

Malin Ahlberg, Markus Forsberg, and Mans Hulden. 2015. Paradigm classification in supervised learning of morphology. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1024–1029, Denver, Colorado. Association for Computational Linguistics.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Khuyagbaatar Batsuren, Omer Goldman, Salam Khalifa, Nizar Habash, Witold Kieraś, Gábor Bella, Brian Leonard, Garrett Nicolai, Kyle Gorman, Yustinus Ghanggo Ate, Maria Ryskina, Sabrina Mielke, Elena Budianskaya, Charbel El-Khaissi, Tiago Pimentel, Michael Gasser, William Abbott Lane, Mohit Raj, Matt Coler, Jaime Rafael Montoya Samame, Delio Siticonatzi Camaiteri, Esaú Zumaeta Rojas, Didier López Francis, Arturo Oncevay, Juan López Bautista, Gema Celeste Silva Villegas, Lucas Torroba Hennigen, Adam Ek, David Guriel, Peter Dirix, Jean-Philippe Bernardy, Andrey Scherbakov, Aziyana Bayyr-ool, Antonios Anastasopoulos, Roberto Zariquiey, Karina Sheifer, Sofya Ganieva, Hilaria Cruz, Ritván Karahóǧa, Stella Markantonatou, George Pavlidis, Matvey Plugaryov, Elena Klyachko, Ali Salehi, Candy Angulo, Jatayu Baxi, Andrew Krizhanovsky, Natalia Krizhanovskaya, Elizabeth Salesky, Clara Vania, Sardana Ivanova, Jennifer White, Rowan Hall Maudslay, Josef Valvoda, Ran Zmigrod, Paula Czarnowska, Irene Nikkarinen, Aelita Salchak, Brijesh Bhatt, Christopher Straughn, Zoey Liu, Jonathan North Washington, Yuval Pinter, Duygu Ataman, Marcin Wolinski, Totok Suhardijanto, Anna Yablonskaya, Niklas Stoehr, Hossep Dolatian, Zahroh Nuriah, Shyam Ratan, Francis M. Tyers, Edoardo M. Ponti, Grant Aiton, Aryaman Arora, Richard J. Hatcher, Ritesh Kumar, Jeremiah Young, Daria Rodionova, Anastasia Yemelina, Taras Andrushko, Igor Marchenko, Polina Mashkovtseva, Alexandra Serova, Emily Prud'hommeaux, Maria Nepomniashchaya, Fausto Giunchiglia, Eleanor Chodroff, Mans Hulden, Miikka Silfverberg, Arya D. McCarthy, David Yarowsky, Ryan Cotterell, Reut Tsarfaty, and Ekaterina Vylomova. 2022. UniMorph 4.0: Universal Morphology. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 840–855, Marseille, France. European Language Resources Association.

Tadeusz Caliński and Jerzy Harabasz. 1974. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, 3(1):1–27.

Ryan Cotterell, Christo Kirov, Mans Hulden, and Jason Eisner. 2019. On the complexity and typology of inflectional morphological systems. *Transactions of the Association for Computational Linguistics*, 7:327–342.

Ryan Cotterell, Christo Kirov, John Sylak-Glassman, David Yarowsky, Jason Eisner, and Mans Hulden. 2016. The sigmorphon 2016 shared task—morphological reinflection. In *Proceedings of the 14th SIGMORPHON workshop on computational research in phonetics, phonology, and morphology*, pages 10–22.

Ryan Cotterell, John Sylak-Glassman, and Christo Kirov. 2017. Neural graphical models over strings for principal parts morphological paradigm completion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 759–765, Valencia, Spain. Association for Computational Linguistics.

Greg Durrett and John DeNero. 2013. Supervised learning of complete morphological paradigms. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1185–1195, Atlanta, Georgia. Association for Computational Linguistics.

Manaal Faruqui, Yulia Tsvetkov, Graham Neubig, and Chris Dyer. 2016. Morphological inflection generation using character sequence to sequence learning. *arXiv preprint arXiv:1512.06110*.

Raphael Finkel and Gregory Stump. 2007. Principal parts and morphological typology. *Morphology*, 17:39–75.

Brendan J. Frey and Dmitri Dueck. 2007. Clustering by passing messages between data points. *Science*, 315(5814):972–976.

Omer Goldman, Khuyagbaatar Batsuren, Salam Khalifa, Aryaman Arora, Garrett Nicolai, Reut Tsarfaty, and Ekaterina Vylomova. 2023. SIGMORPHON–UniMorph 2023 shared task 0: Typologically diverse morphological inflection. In *Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 117–125, Toronto, Canada. Association for Computational Linguistics.

Omer Goldman, David Guriel, and Reut Tsarfaty. 2021. (un) solving morphological inflection: Lemma overlap artificially inflates models' performance. *arXiv preprint arXiv:2108.05682*.

Mans Hulden. 2014. Generalizing inflection tables into paradigms with finite state operations. In *Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM*, pages 29–36, Baltimore, Maryland. Association for Computational Linguistics.

Katharina Kann, Ryan Cotterell, and Hinrich Schütze. 2017. Neural multi-source morphological reinflection. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 514–524, Valencia, Spain. Association for Computational Linguistics.

Vladimir I. Levenshtein. 1966. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet physics doklady*, 10(8):707–710.

Ling Liu and Mans Hulden. 2020. Leveraging principal parts for morphological inflection. In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 153–161.

Eugene W Myers. 1986. An o (nd) difference algorithm and its variations. *Algorithmica*, 1(1):251–266.

Garrett Nicolai, Colin Cherry, and Grzegorz Kondrak. 2015. Inflection generation as discriminative string transduction. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 922–931, Denver, Colorado. Association for Computational Linguistics.

Gregory Stump and Raphael A. Finkel. 2013. *Principal parts*, Cambridge Studies in Linguistics, page 9–39. Cambridge University Press.

Robert A Wagner and Michael J Fischer. 1974. The string-to-string correction problem. *Journal of the ACM (JACM)*, 21(1):168–173.

Appendix

A Technical Overview of the PRINCIPAL PARTS DETECTION Dataset

This section provides the technical details of the PRINCIPAL PARTS DETECTION dataset, including the number of samples per feature set in each language's verb paradigm and the total number of gold principal parts for each language. In some cases, specific feature sets were removed for various reasons, which are explained in subsection A.2.

Additionally, we list the gold principal parts for each language, formatted as `feature_set (e.g., form)`. When two feature sets share the same form, the gold principal parts are listed in square brackets `[ ]`. The first feature set corresponds to the principal part identified in linguistic literature, while the second represents a feature set that consistently shares the same form across all samples in the dataset. In such cases, the second feature set is included as a possible principal part, as the algorithm's choice between them does not affect the analysis. To avoid redundancy, no principal part is counted more than once in these scenarios.

A.1 Dataset Summary and Illustrative Lexeme Examples

For each language, we provide an example lexeme to illustrate the principal parts, formatted as `feature_set (e.g., form)`. These examples are illustrative and may not share the same meanings across languages.

A.2 Explanatory Notes

The following explanatory notes clarify decisions made during dataset preparation and supplement the information presented in Table 6:

- **Spanish**: PRO feature sets, representing verbs with object clitic pronouns, were removed.
- **Swedish**: The V-IMP-PASS feature set was excluded due to insufficient samples (only three).
- **Latin**:
  - Passive feature sets were excluded.
  - Feature sets starting with V.PTCP (instead of V-V.PTCP) were removed.

Language	Features	Samples per Feature Set	# of Gold Principal Parts	Gold Principal Parts
English	5	23,896–31,848	3	V-NFIN-IMP+SBJV (e.g., eat), V-PST (e.g., ate), V-V.PTCP-PST (e.g., eaten)
French	49	7,483–7,535	7	V-NFIN (e.g., mangier), V-IND-PRS-1-PL (e.g., manjons), V.PTCP-PST (e.g., mangié), V-IND-FUT-1-SG (e.g., mangerai), V-IND-PRS-1-SG (e.g., manju), V-IND-PRS-3-PL (e.g., manjüent), V-IND-PST-1-SG-PFV (e.g., manjai)
German	30	2,307–6,661	5	V-NFIN (e.g., essen), V.PTCP-PST (e.g., gegessen), [V-IND-SG-3-PST, V-IND-SG-1-PST (e.g., aß)], V-IND-SG-3-PRS (e.g., isst), [V-SBJV-SG-3-PST, V-SBJV-SG-1-PST (e.g., äße)]
Spanish	79	6,676–6,695	2	V-NFIN (e.g., comer), V-IND-PRS-1-SG (e.g., como)
Danish	8	162	5	V-ACT-NFIN (e.g., danse), V-ACT-IND-PRS (e.g., danser), V-ACT-IND-PST (e.g., dansede), V-ACT-IMP (e.g., dans), V.PTCP-PASS-PST (e.g., danset)
Swedish	19	2,114–2,536	5	[V-NFIN-ACT, V-IND-PL-ACT-PRS (e.g., äta)], V-IND-SG-ACT-PRS (e.g., äter), V-IND-SG-ACT-PST (e.g., åt), V-V.CVB-ACT (e.g., ätit), V-IMP-ACT (e.g., ät)
Finnish	161	7,221–7,226	4	V-NFIN-ACT+PASS (e.g., syödä), V-ACT-PRS-POS-IND-1-SG (e.g., syön), V-ACT-PST-POS-IND-3-SG (e.g., söi), V.PTCP-ACT-PST (e.g., syönyt)
Turkish	703	588	2	V-NFIN (e.g., içmek), V-IND-PRS-HAB-3-SG-POS-DECL (e.g., içer)
Latin	48	450–947	3	V-IND-ACT-PRS-1-SG (e.g., -pleō), V-NFIN-ACT-PRS (e.g., -plēre), V-V.MSDR-ACC-LGSPEC1 (e.g., -plētum)

Table 6: Summary of the PRINCIPAL PARTS DETECTION dataset by language, including gold principal parts and illustrative lexeme examples.

– Feature sets with 30 or fewer samples were excluded.

– The first-person-singular-perfect-active-indicative feature set was excluded from the gold principal parts list due to insufficient data (only two samples).

B Characterization Tables for Selected Languages

To illustrate the structure of the characterization methods, we present detailed characterization tables for three representative languages from our dataset. These tables demonstrate how different feature sets relate within their verb paradigms, showcasing the variation across Edit Distance, Edit Script, and Reinflection Accuracy characterization methods.

Each language is represented by three tables, corresponding to the distinct characterization methods, with principal parts highlighted in yellow for clarity. Additionally, cases where two feature sets consistently share the same form and are interchangeable as principal parts are marked with a distinct color. Since these feature sets carry identical information, the model's selection between them does not impact the results.

Interpretation of Tables. The provided tables exemplify the structure of the characterization methods rather than an exhaustive display of all ten languages in our study. While specific lexeme examples are shown in the rows and columns, the quantified relationships they capture apply to the entire verb paradigm of each language. These examples serve to illustrate the broader implicative patterns identified during the characterization process.

B.1 Characterization Tables for English

B.2 Characterization Tables for German

B.3 Characterization Tables for Swedish

	Features	V-NFIN-IMP+SBJV - eat	V-PRS-3-SG - eats	V-PST - ate	V-V.PTCP-PRS - eating	V-V.PTCP-PST - eaten
1	V-NFIN-IMP+SBJV - eat	0	1.157683294	1.532683294	3.088508537	1.534943087
2	V-PRS-3-SG - eats	1.157683294	0	1.421493137	3.087504185	1.410905591
3	V-PST - ate	1.532683294	1.421493137	0	3.066078005	0.048627385
4	V-V.PTCP-PRS - eating	3.088508537	3.087504185	3.066078005	0	3.034273519
5	V-V.PTCP-PST - eaten	1.534943087	1.410905591	0.048627385	3.034273519	0

Figure 1: Average edit distances for the English verb paradigm. Values range from 0 to 4.5. Darker red shades indicate closer relationships between feature sets, while darker turquoise shades represent greater differences.

	Features	V-NFIN-IMP+SBJV - eat	V-PRS-3-SG - eats	V-PST - ate	V-V.PTCP-PRS - eating	V-V.PTCP-PST - eaten
1	V-NFIN-IMP+SBJV - eat	1	27	117	51	124
2	V-PRS-3-SG - eats	29	1	110	48	117
3	V-PST - ate	124	110	1	116	43
4	V-V.PTCP-PRS - eating	55	59	119	1	121
5	V-V.PTCP-PST - eaten	128	118	45	119	1

Figure 2: Edit Script scores for the English verb paradigm. Values range from 1 to 128. Darker purple shades indicate fewer unique character sets (closer relationships), while darker air-force-blue shades reflect greater variation.

	Features	V-NFIN-IMP+SBJV - eat	V-PRS-3-SG - eats	V-PST - ate	V-V.PTCP-PRS - eating	V-V.PTCP-PST - eaten
1	V-NFIN-IMP+SBJV - eat	0.95	0.96	0.92	0.94	0.92
2	V-PRS-3-SG - eats	0.95	0.96	0.92	0.94	0.91
3	V-PST - ate	0.9	0.91	0.96	0.94	0.95
4	V-V.PTCP-PRS - eating	0.91	0.92	0.92	0.95	0.92
5	V-V.PTCP-PST - eaten	0.91	0.91	0.96	0.95	0.96

Figure 3: Reinflection Accuracy scores for the English verb paradigm. Values range from 0.9 to 0.96. Darker teal shades indicate higher accuracy, while darker pink shades reflect lower performance.

	Features	V.PTCP-PRS - essend	V.PTCP-PST - gegessen	V-IMP-PL-2 - esst	V-IMP-SG-2 - iss	V-IND-PL-1-PRS - essen	V-IND-PL-1-PST - aßen	V-IND-PL-2-PRS - esst	V-IND-PL-2-PST - aßt	V-IND-PL-3-PRS - essen	V-IND-PL-3-PST - aßen	V-IND-SG-1-PRS - esse	V-IND-SG-1-PST - aß	V-IND-SG-2-PRS - isst	V-IND-SG-2-PST - aßt	V-IND-SG-3-PRS - isst
1	V.PTCP-PRS - essend	0	3.828404144	4.705590622	4.09137361	3.367172248	4.662561946	4.694548731	5.38954798	3.367172248	4.662561946	4.039945938	4.934224358	5.010211744	5.82640036	4.760774891
2	V.PTCP-PST - gegessen	3.828404144	0	4.06372107	4.592425609	5.027481604	5.257396005	4.052710617	5.421985283	5.027481604	5.257396005	4.559393302	4.620964109	4.822345698	6.212794714	4.093707764
3	V-IMP-PL-2 - esst	4.705590622	4.06372107	0	1.084310189	1.755034566	2.345807033	0.019687406	2.015479411	1.755034566	2.345807033	1.04854223	1.381574992	1.011722272	2.960925759	0.129996994
4	V-IMP-SG-2 - iss	4.09137361	4.592425609	1.084310189	0	1.131650135	2.366847009	1.093477607	2.331680192	1.131650135	2.366847009	0.154192967	1.460324617	1.988578299	3.282837391	1.086113616
5	V-IND-PL-1-PRS - essen	3.367172248	5.027481604	1.755034566	1.131650135	0	1.42183511	1.740201231	2.560444511	0	1.42183511	1.038594384	2.316714221	2.094158282	3.415227512	1.845171948
6	V-IND-PL-1-PST - aßen	4.662561946	5.257396005	2.345807033	2.366847009	1.42183511	0	2.333083046	1.159032888	1.42183511	0	2.313710767	1.191169845	3.136206638	2.000450518	2.363417931
7	V-IND-PL-2-PRS - esst	4.694548731	4.052710617	0.019687406	1.093477607	1.740201231	2.333083046	0	2.003003454	1.740201231	2.333083046	1.036642138	1.36672173	0.997447064	2.948490764	0.118786605
8	V-IND-PL-2-PST - aßt	5.38954798	5.421985283	2.015479411	2.331680192	2.560444511	1.159032888	2.003003454	0	2.560444511	1.159032888	2.313110077	1.031986785	2.158282024	0.970866496	2.043700255
9	V-IND-PL-3-PRS - essen	3.367172248	5.027481604	1.755034566	1.131650135	0	1.42183511	1.740201231	2.560444511	0	1.42183511	1.038594384	2.316714221	2.094158282	3.415227512	1.845171948
10	V-IND-PL-3-PST - aßen	4.662561946	5.257396005	2.345807033	2.366847009	1.42183511	0	2.333083046	1.159032888	1.42183511	0	2.313710767	1.191169845	3.136206638	2.000450518	2.363417931
11	V-IND-SG-1-PRS - esse	4.039945938	4.559393302	1.04854223	0.154192967	1.038594384	2.313710767	1.036642138	2.313110077	1.038594384	2.313710767	0	1.475747109	2.021925214	3.263402913	1.116083496
12	V-IND-SG-1-PST - aß	4.934224358	4.620964109	1.381574992	1.460324617	2.316714221	1.191169845	1.36672173	1.031986785	2.316714221	1.191169845	1.475747109	0	2.272413275	2.002853281	1.36506983
13	V-IND-SG-2-PRS - isst	5.010211744	4.822345698	1.011722272	1.988578299	2.094158282	3.136206638	0.997447064	2.158282024	2.094158282	3.136206638	2.021925214	2.272413275	0	2.094008109	0.921009161
14	V-IND-SG-2-PST - aßt	5.82640036	6.212794714	2.960925759	3.282837391	3.415227512	2.000450518	2.948490764	0.970866496	3.415227512	2.000450518	3.263402913	2.002853281	2.094008109	0	2.988436702
15	V-IND-SG-3-PRS - isst	4.760774891	4.093707764	0.129996994	1.086113616	1.845171948	2.363417931	0.118786605	2.043700255	1.845171948	2.363417931	1.116083496	1.36506983	0.921009161	2.988436702	0
16	V-IND-SG-3-PST - aß	4.934224358	4.620964109	1.381574992	1.460324617	2.316714221	1.191169845	1.36672173	1.031986785	2.316714221	1.191169845	1.475747109	0	2.272413275	2.002853281	1.36506983
17	V-NFIN - essen	0.999849872	2.839963969	3.906672678	3.317703637	2.669019372	3.942033338	3.894728938	4.772488362	2.669019372	3.942033338	3.270761376	4.30830455	4.393001952	5.633428443	3.951944736
18	V-NFIN-LGSPEC01 - אין	3.000433463	3.605981795	9.666521928	9.226226661	9.286767896	10.21822126	9.650759219	10.52494577	9.286767896	10.21822126	9.195227766	9.907158351	10.22212581	11.08286334	9.709761388
19	V-SBJV-PL-1-PRS - essen	3.367172248	5.027481604	1.755034566	1.131650135	0	1.42183511	1.740201231	2.560444511	0	1.42183511	1.038594384	2.316714221	2.094158282	3.415227512	1.845171948
20	V-SBJV-PL-1-PST - äßen	4.664664364	5.299294188	2.345807033	2.366245867	1.419132002	0.133503529	2.331280973	1.290734344	1.419132002	0.133503529	2.313260249	1.322420784	3.135605947	2.133053011	2.364018621
21	V-SBJV-PL-2-PRS - esset	4.326325274	4.607748911	0.919446949	1.111361587	1.161435651	2.208890224	0.901787055	1.443760324	1.161435651	2.208890224	1.059168043	2.234419583	1.117284878	2.297492116	1.019522451
22	V-SBJV-PL-2-PST - äßet	5.367772939	5.395404715	2.154042681	2.363540727	2.418831656	1.133503529	2.13966061	0.290734344	2.418831656	1.133503529	2.313110077	1.322120439	2.200030035	1.133053011	2.181859138
23	V-SBJV-PL-3-PRS - essen	3.367172248	5.027481604	1.755034566	1.131650135	0	1.42183511	1.740201231	2.560444511	0	1.42183511	1.038594384	2.316714221	2.094158282	3.415227512	1.845171948
24	V-SBJV-PL-3-PST - äßen	4.664664364	5.299294188	2.345807033	2.366245867	1.419132002	0.133503529	2.331280973	1.290734344	1.419132002	0.133503529	2.313260249	1.322420784	3.135605947	2.133053011	2.364018621
25	V-SBJV-SG-1-PRS - esse	4.023126596	4.546628623	1.039074241	0.117222723	1.020874005	2.302297642	1.022075387	2.300946088	1.020874005	2.302297642	0.063823397	1.463132602	2.015768133	3.251689443	1.11353056
26	V-SBJV-SG-1-PST - äße	4.908394654	4.591680432	1.393748121	1.373609859	2.184562247	1.132902838	1.377083646	1.132602493	2.184562247	1.132902838	1.316413876	0.322570957	2.277819492	2.101066226	1.395855234
27	V-SBJV-SG-2-PRS - essest	4.742003304	5.460279321	1.916290953	2.108055305	2.160684788	3.206337288	1.900886019	2.304099715	2.160684788	3.206337288	2.055263553	3.211292987	1.099564499	1.450518096	2.016068479
28	V-SBJV-SG-2-PST - äßest	5.826700706	6.24598288	3.101893598	3.349113315	3.418381138	2.132602493	3.087550683	1.290734344	3.418381138	2.132602493	3.302297642	2.321970266	2.24207839	0.320768884	3.123291786
29	V-SBJV-SG-3-PRS - esse	4.023126596	4.546628623	1.039074241	0.117222723	1.020874005	2.302297642	1.022075387	2.300946088	1.020874005	2.302297642	0.063823397	1.463132602	2.015768133	3.251689443	1.11353056
30	V-SBJV-SG-3-PST - äße	4.908394654	4.591680432	1.393748121	1.373609859	2.184562247	1.132902838	1.377083646	1.132602493	2.184562247	1.132902838	1.316413876	0.322570957	2.277819492	2.101066226	1.395855234

	Features	V-IND-SG-3-PST - aß	V-NFIN - essen	V-NFIN-LGSPEC01 - אין	V-SBJV-PL-1-PRS - essen	V-SBJV-PL-1-PST - äßen	V-SBJV-PL-2-PRS - esset	V-SBJV-PL-2-PST - äßet	V-SBJV-PL-3-PRS - essen	V-SBJV-PL-3-PST - äßen	V-SBJV-SG-1-PRS - esse	V-SBJV-SG-1-PST - äße	V-SBJV-SG-2-PRS - essest	V-SBJV-SG-2-PST - äßest	V-SBJV-SG-3-PRS - esse	V-SBJV-SG-3-PST - äße
1	V.PTCP-PRS - essend	4.934224358	0.999849872	3.000433463	3.367172248	4.664664364	4.326325274	5.367772939	3.367172248	4.664664364	4.023126596	4.908394654	4.742003304	5.826700706	4.023126596	4.908394654
2	V.PTCP-PST - gegessen	4.620964109	2.839963969	3.605981795	5.027481604	5.299294188	4.607748911	5.395404715	5.027481604	5.299294188	4.546628623	4.591680432	5.460279321	6.24598288	4.546628623	4.591680432
3	V-IMP-PL-2 - esst	1.381574992	3.906672678	9.666521928	1.755034566	2.345807033	0.919446949	2.154042681	1.755034566	2.345807033	1.039074241	1.393748121	1.916290953	3.101893598	1.039074241	1.393748121
4	V-IMP-SG-2 - iss	1.460324617	3.317703637	9.226226661	1.131650135	2.366245867	1.111361587	2.363540727	1.131650135	2.366245867	0.117222723	1.373609859	2.108055305	3.349113315	0.117222723	1.373609859
5	V-IND-PL-1-PRS - essen	2.316714221	2.669019372	9.286767896	0	1.419132002	1.161435651	2.418831656	0	1.419132002	1.020874005	2.184562247	2.160684788	3.418381138	1.020874005	2.184562247
6	V-IND-PL-1-PST - aßen	1.191169845	3.942033338	10.21822126	1.42183511	0.133503529	2.208890224	1.133503529	1.42183511	0.133503529	2.302297642	1.132902838	3.206337288	2.132602493	2.302297642	1.132902838
7	V-IND-PL-2-PRS - esst	1.36672173	3.894728938	9.650759219	1.740201231	2.331280973	0.901787055	2.13966061	1.740201231	2.331280973	1.022075387	1.377083646	1.900886019	3.087550683	1.022075387	1.377083646
8	V-IND-PL-2-PST - aßt	1.031986785	4.772488362	10.52494577	2.560444511	1.290734344	1.443760324	0.290734344	2.560444511	1.290734344	2.300946088	1.132602493	2.304099715	1.290734344	2.300946088	1.132602493
9	V-IND-PL-3-PRS - essen	2.316714221	2.669019372	9.286767896	0	1.419132002	1.161435651	2.418831656	0	1.419132002	1.020874005	2.184562247	2.160684788	3.418381138	1.020874005	2.184562247
10	V-IND-PL-3-PST - aßen	1.191169845	3.942033338	10.21822126	1.42183511	0.133503529	2.208890224	1.133503529	1.42183511	0.133503529	2.302297642	1.132902838	3.206337288	2.132602493	2.302297642	1.132902838
11	V-IND-SG-1-PRS - esse	1.475747109	3.270761376	9.195227766	1.038594384	2.313260249	1.059168043	2.313110077	1.038594384	2.313260249	0.063823397	1.316413876	2.055263553	3.302297642	0.063823397	1.316413876
12	V-IND-SG-1-PST - aß	0	4.30830455	9.907158351	2.316714221	1.322420784	2.234419583	1.322120439	2.316714221	1.322420784	1.463132602	0.322570957	3.211292987	2.321970266	1.463132602	0.322570957
13	V-IND-SG-2-PRS - isst	2.272413275	4.393001952	10.22212581	2.094158282	3.135605947	1.117284878	2.200030035	2.094158282	3.135605947	2.015768133	2.277819492	1.099564499	2.24207839	2.015768133	2.277819492
14	V-IND-SG-2-PST - aßt	2.002853281	5.633428443	11.08286334	3.415227512	2.133053011	2.297492116	1.133053011	3.415227512	2.133053011	3.251689443	2.101066226	1.450518096	0.320768884	3.251689443	2.101066226
15	V-IND-SG-3-PRS - isst	1.36506983	3.951944736	9.709761388	1.845171948	2.364018621	1.019522451	2.181859138	1.845171948	2.364018621	1.11353056	1.395855234	2.016068479	3.123291786	1.11353056	1.395855234
16	V-IND-SG-3-PST - aß	0	4.30830455	9.907158351	2.316714221	1.322420784	2.234419583	1.322120439	2.316714221	1.322420784	1.463132602	0.322570957	3.211292987	2.321970266	1.463132602	0.322570957
17	V-NFIN - essen	4.30830455	0	2.000433463	2.669019372	3.945337138	3.671572308	4.759723682	2.669019372	3.945337138	3.253791861	4.291184863	4.622916354	5.702057366	3.253791861	4.291184863
18	V-NFIN-LGSPEC01 - אין	9.907158351	2.000433463	0	9.286767896	10.23991323	9.90021692	10.5813449	9.286767896	10.23991323	9.161388286	9.966160521	10.52277657	11.19392625	9.161388286	9.966160521
19	V-SBJV-PL-1-PRS - essen	2.316714221	2.669019372	9.286767896	0	1.419132002	1.161435651	2.418831656	0	1.419132002	1.020874005	2.184562247	2.160684788	3.418381138	1.020874005	2.184562247
20	V-SBJV-PL-1-PST - äßen	1.322420784	3.945337138	10.23991323	1.419132002	0	2.206187115	1.000300345	1.419132002	0	2.299594534	1.000300345	3.203934525	2.000300345	2.299594534	1.000300345
21	V-SBJV-PL-2-PRS - esset	2.234419583	3.671572308	9.90021692	1.161435651	2.206187115	0	1.302447815	1.161435651	2.206187115	1	2.109475897	1	2.300795915	1	2.109475897
22	V-SBJV-PL-2-PST - äßet	1.322120439	4.759723682	10.5813449	2.418831656	1.000300345	1.302447815	0	2.418831656	1.000300345	2.299294188	1	2.20588677	1	2.299294188	1
23	V-SBJV-PL-3-PRS - essen	2.316714221	2.669019372	9.286767896	0	1.419132002	1.161435651	2.418831656	0	1.419132002	1.020874005	2.184562247	2.160684788	3.418381138	1.020874005	2.184562247
24	V-SBJV-PL-3-PST - äßen	1.322420784	3.945337138	10.23991323	1.419132002	0	2.206187115	1.000300345	1.419132002	0	2.299594534	1.000300345	3.203934525	2.000300345	2.299594534	1.000300345
25	V-SBJV-SG-1-PRS - esse	1.463132602	3.253791861	9.161388286	1.020874005	2.299594534	1	2.299294188	1.020874005	2.299594534	0	1.301546779	2	3.288331581	0	1.301546779
26	V-SBJV-SG-1-PST - äße	0.322570957	4.291184863	9.966160521	2.184562247	1.000300345	2.109475897	1	2.184562247	1.000300345	1.301546779	0	3.099864845	2	1.301546779	0
27	V-SBJV-SG-2-PRS - essest	3.211292987	4.622916354	10.52277657	2.160684788	3.203934525	1	2.20588677	2.160684788	3.203934525	2	3.099864845	0	1.303649197	2	3.099864845
28	V-SBJV-SG-2-PST - äßest	2.321970266	5.702057366	11.19392625	3.418381138	2.000300345	2.300795915	1	3.418381138	2.000300345	3.288331581	2	1.303649197	0	3.288331581	2
29	V-SBJV-SG-3-PRS - esse	1.463132602	3.253791861	9.161388286	1.020874005	2.299594534	1	2.299294188	1.020874005	2.299594534	0	1.301546779	2	3.288331581	0	1.301546779
30	V-SBJV-SG-3-PST - äße	0.322570957	4.291184863	9.966160521	2.184562247	1.000300345	2.109475897	1	2.184562247	1.000300345	1.301546779	0	3.099864845	2	1.301546779	0

Figure 4: Average edit distances for the German verb paradigm. Values range from 0 to 11.19. Darker red shades indicate closer relationships between feature sets, while darker ball-blue shades represent greater distances.

	Features	V.PTCP-PRS - essend	V.PTCP-PST - gegessen	V-IMP-PL-2 - esst	V-IMP-SG-2 - iss	V-IND-PL-1-PRS - essen	V-IND-PL-1-PST - aßen	V-IND-PL-2-PRS - esst	V-IND-PL-2-PST - aßt	V-IND-PL-3-PRS - essen	V-IND-PL-3-PST - aßen	V-IND-SG-1-PRS - esse	V-IND-SG-1-PST - aß	V-IND-SG-2-PRS - isst	V-IND-SG-2-PST - aßt	V-IND-SG-3-PRS - isst	V-IND-SG-3-PST - aß	V-NFIN - essen	V-NFIN-LGSPEC01 - אין	V-SBJV-PL-1-PRS - essen	V-SBJV-PL-1-PST - äßen	V-SBJV-PL-2-PRS - esset	V-SBJV-PL-2-PST - äßet	V-SBJV-PL-3-PRS - essen	V-SBJV-PL-3-PST - äßen	V-SBJV-SG-1-PRS - esse	V-SBJV-SG-1-PST - äße	V-SBJV-SG-2-PRS - essest	V-SBJV-SG-2-PST - äßest	V-SBJV-SG-3-PRS - esse	V-SBJV-SG-3-PST - äße
1	V.PTCP-PRS - essend	1	141	473	511	352	892	466	876	352	892	417	847	652	913	586	847	2	6	352	896	405	854	352	896	400	849	421	869	400	849
2	V.PTCP-PST - gegessen	102	1	885	1026	982	845	884	880	982	845	963	810	1010	947	926	810	101	81	982	926	900	931	982	926	938	901	943	974	938	901
3	V-IMP-PL-2 - esst	551	922	1	32	12	69	11	64	12	69	29	70	28	77	25	70	559	579	12	80	18	81	12	80	22	80	19	91	22	80
4	V-IMP-SG-2 - iss	554	1107	37	1	39	108	45	122	39	108	41	118	50	135	45	118	561	564	39	117	31	118	39	117	30	116	33	119	30	116
5	V-IND-PL-1-PRS - essen	404	1107	13	33	1	58	3	74	1	58	24	69	25	86	19	69	412	422	1	65	9	64	1	65	9	64	9	64	9	64
6	V-IND-PL-1-PST - aßen	912	924	83	110	58	1	77	3	58	1	85	4	89	5	81	4	909	886	58	15	72	15	58	15	72	15	72	15	72	15
7	V-IND-PL-2-PRS - esst	544	921	11	40	2	63	1	58	2	63	24	64	23	72	21	64	552	581	2	72	8	73	2	72	12	72	9	83	12	72
8	V-IND-PL-2-PST - aßt	860	912	65	111	64	2	59	1	64	2	90	4	80	4	70	4	857	844	64	19	78	19	64	19	78	19	85	23	78	19
9	V-IND-PL-3-PRS - essen	404	1107	13	33	1	58	3	74	1	58	24	69	25	86	19	69	412	422	1	65	9	64	1	65	9	64	9	64	9	64
10	V-IND-PL-3-PST - aßen	912	924	83	110	58	1	77	3	58	1	85	4	89	5	81	4	909	886	58	15	72	15	58	15	72	15	72	15	72	15
11	V-IND-SG-1-PRS - esse	469	1066	38	40	29	86	33	101	29	86	1	96	44	114	38	96	482	489	29	96	31	95	29	96	34	95	31	95	34	95
12	V-IND-SG-1-PST - aß	833	832	69	104	58	4	63	4	58	4	83	1	79	9	68	1	825	816	58	21	75	23	58	21	71	20	82	28	71	20
13	V-IND-SG-2-PRS - isst	715	1029	28	49	25	81	23	79	25	81	42	81	1	76	6	81	719	731	25	91	34	89	25	91	39	91	38	96	39	91
14	V-IND-SG-2-PST - aßt	904	985	78	125	79	5	73	4	79	5	106	9	76	1	85	9	901	877	79	25	90	24	79	25	93	25	94	27	93	25
15	V-IND-SG-3-PRS - isst	643	949	25	41	18	69	21	70	18	69	32	68	6	85	1	68	648	666	18	78	28	81	18	78	29	78	32	90	29	78
16	V-IND-SG-3-PST - aß	833	832	69	104	58	4	63	4	58	4	83	1	79	9	68	1	825	816	58	21	75	23	58	21	71	20	82	28	71	20
17	V-NFIN - essen	2	138	485	519	361	888	478	878	361	888	427	842	657	915	590	842	1	6	361	892	415	854	361	892	410	849	432	869	410	849
18	V-NFIN-LGSPEC01 - אין	6	103	490	512	370	851	492	831	370	851	426	802	654	860	588	802	6	1	370	853	429	821	370	853	422	815	445	835	422	815
19	V-SBJV-PL-1-PRS - essen	404	1107	13	33	1	58	3	74	1	58	24	69	25	86	19	69	412	422	1	65	9	64	1	65	9	64	9	64	9	64
20	V-SBJV-PL-1-PST - äßen	881	984	94	119	65	15	86	20	65	15	95	21	99	25	90	21	879	851	65	1	79	2	65	1	79	2	79	2	79	2
21	V-SBJV-PL-2-PRS - esset	479	986	22	30	13	74	12	88	13	74	31	87	35	100	31	87	490	510	13	81	1	79	13	81	1	81	1	79	1	81
22	V-SBJV-PL-2-PST - äßet	853	976	94	120	65	15	86	19	65	15	95	23	100	24	92	23	853	832	65	2	79	1	65	2	79	1	79	1	79	1
23	V-SBJV-PL-3-PRS - essen	404	1107	13	33	1	58	3	74	1	58	24	69	25	86	19	69	412	422	1	65	9	64	1	65	9	64	9	64	9	64
24	V-SBJV-PL-3-PST - äßen	881	984	94	119	65	15	86	20	65	15	95	21	99	25	90	21	879	851	65	1	79	2	65	1	79	2	79	2	79	2
25	V-SBJV-SG-1-PRS - esse	451	1046	26	29	13	74	16	89	13	74	33	84	39	101	33	84	462	482	13	81	1	81	13	81	1	81	1	81	1	81
26	V-SBJV-SG-1-PST - äße	846	931	93	118	65	15	85	19	65	15	95	20	98	24	89	20	846	823	65	2	79	1	65	2	79	1	79	1	79	1
27	V-SBJV-SG-2-PRS - essest	508	1049	23	32	13	78	13	92	13	78	31	96	37	101	35	96	520	544	13	84	1	79	13	84	1	84	1	79	1	84
28	V-SBJV-SG-2-PST - äßest	886	1044	104	121	68	15	96	23	68	15	98	28	107	27	101	28	885	862	68	2	79	1	68	2	82	1	80	1	82	1
29	V-SBJV-SG-3-PRS - esse	451	1046	26	29	13	74	16	89	13	74	33	84	39	101	33	84	462	482	13	81	1	81	13	81	1	81	1	81	1	81
30	V-SBJV-SG-3-PST - äße	846	931	93	118	65	15	85	19	65	15	95	20	98	24	89	20	846	823	65	2	79	1	65	2	79	1	79	1	79	1

Figure 5: Edit Script scores for the German Verb Paradigm. Values range from 1 to 1,107. Darker purple shades indicate fewer unique character sets (closer relationships), while darker air-force-blue shades reflect greater variation.

	Features	V.PTCP-PRS - essend	V.PTCP-PST - gegessen	V-IMP-PL-2 - esst	V-IMP-SG-2 - iss	V-IND-PL-1-PRS - essen	V-IND-PL-1-PST - aßen	V-IND-PL-2-PRS - esst	V-IND-PL-2-PST - aßt	V-IND-PL-3-PRS - essen	V-IND-PL-3-PST - aßen	V-IND-SG-1-PRS - esse	V-IND-SG-1-PST - aß	V-IND-SG-2-PRS - isst	V-IND-SG-2-PST - aßt	V-IND-SG-3-PRS - isst	V-IND-SG-3-PST - aß	V-NFIN - essen	V-NFIN-LGSPEC01 - אין	V-SBJV-PL-1-PRS - essen	V-SBJV-PL-1-PST - äßen	V-SBJV-PL-2-PRS - esset	V-SBJV-PL-2-PST - äßet	V-SBJV-PL-3-PRS - essen	V-SBJV-PL-3-PST - äßen	V-SBJV-SG-1-PRS - esse	V-SBJV-SG-1-PST - äße	V-SBJV-SG-2-PRS - essest	V-SBJV-SG-2-PST - äßest	V-SBJV-SG-3-PRS - esse	V-SBJV-SG-3-PST - äße
1	V.PTCP-PRS - essend	0.85	0.75	0.82	0.8	0.82	0.77	0.82	0.77	0.83	0.77	0.81	0.78	0.78	0.77	0.79	0.78	0.85	0.81	0.82	0.77	0.83	0.77	0.83	0.77	0.83	0.77	0.83	0.77	0.83	0.77
2	V.PTCP-PST - gegessen	0.83	0.87	0.83	0.83	0.84	0.85	0.83	0.85	0.85	0.85	0.82	0.84	0.83	0.84	0.83	0.85	0.83	0.81	0.85	0.84	0.85	0.84	0.85	0.84	0.85	0.84	0.85	0.84	0.85	0.84
3	V-IMP-PL-2 - esst	0.86	0.8	0.89	0.86	0.89	0.86	0.89	0.86	0.89	0.86	0.88	0.86	0.86	0.86	0.88	0.86	0.86	0.84	0.88	0.86	0.88	0.84	0.88	0.86	0.89	0.86	0.88	0.85	0.89	0.85
4	V-IMP-SG-2 - iss	0.84	0.81	0.88	0.89	0.88	0.85	0.88	0.84	0.88	0.85	0.88	0.84	0.86	0.84	0.87	0.84	0.85	0.85	0.89	0.84	0.89	0.84	0.89	0.85	0.89	0.84	0.89	0.84	0.89	0.84
5	V-IND-PL-1-PRS - essen	0.85	0.81	0.88	0.86	0.9	0.86	0.88	0.85	0.89	0.86	0.87	0.86	0.85	0.85	0.87	0.86	0.85	0.84	0.89	0.85	0.89	0.85	0.89	0.85	0.88	0.85	0.88	0.85	0.88	0.85
6	V-IND-PL-1-PST - aßen	0.84	0.82	0.88	0.86	0.88	0.9	0.88	0.9	0.88	0.9	0.87	0.9	0.86	0.89	0.88	0.9	0.83	0.8	0.88	0.89	0.88	0.89	0.88	0.89	0.88	0.89	0.88	0.89	0.88	0.89
7	V-IND-PL-2-PRS - esst	0.86	0.81	0.89	0.87	0.89	0.86	0.89	0.86	0.89	0.86	0.88	0.85	0.86	0.86	0.88	0.85	0.85	0.84	0.89	0.85	0.88	0.86	0.89	0.86	0.89	0.85	0.88	0.86	0.89	0.85
8	V-IND-PL-2-PST - aßt	0.84	0.83	0.88	0.86	0.87	0.89	0.88	0.89	0.87	0.89	0.87	0.89	0.87	0.89	0.88	0.89	0.84	0.82	0.87	0.87	0.87	0.88	0.87	0.88	0.87	0.88	0.87	0.88	0.88	0.88
9	V-IND-PL-3-PRS - essen	0.86	0.8	0.88	0.86	0.89	0.85	0.88	0.85	0.88	0.85	0.87	0.85	0.86	0.85	0.87	0.85	0.85	0.84	0.88	0.85	0.88	0.85	0.88	0.85	0.88	0.85	0.88	0.84	0.88	0.85
10	V-IND-PL-3-PST - aßen	0.85	0.83	0.88	0.87	0.88	0.9	0.88	0.9	0.88	0.9	0.88	0.9	0.87	0.9	0.88	0.9	0.86	0.85	0.88	0.89	0.88	0.89	0.88	0.89	0.88	0.89	0.88	0.9	0.88	0.89
11	V-IND-SG-1-PRS - esse	0.85	0.79	0.88	0.86	0.88	0.84	0.87	0.83	0.88	0.84	0.89	0.84	0.84	0.83	0.85	0.84	0.85	0.85	0.88	0.84	0.88	0.84	0.88	0.84	0.88	0.84	0.87	0.84	0.88	0.84
12	V-IND-SG-1-PST - aß	0.84	0.82	0.88	0.86	0.88	0.9	0.88	0.89	0.88	0.9	0.87	0.9	0.86	0.89	0.87	0.89	0.85	0.83	0.88	0.89	0.88	0.88	0.88	0.89	0.88	0.89	0.88	0.89	0.88	0.88
13	V-IND-SG-2-PRS - isst	0.83	0.79	0.87	0.84	0.87	0.85	0.87	0.85	0.87	0.85	0.86	0.85	0.88	0.84	0.87	0.85	0.84	0.84	0.87	0.84	0.87	0.84	0.87	0.85	0.87	0.84	0.86	0.84	0.87	0.84
14	V-IND-SG-2-PST - aßt	0.86	0.85	0.89	0.86	0.89	0.9	0.88	0.9	0.88	0.9	0.87	0.9	0.86	0.9	0.89	0.9	0.86	0.86	0.89	0.9	0.88	0.89	0.89	0.89	0.88	0.89	0.89	0.89	0.88	0.89
15	V-IND-SG-3-PRS - isst	0.84	0.79	0.87	0.86	0.88	0.85	0.87	0.85	0.88	0.85	0.86	0.84	0.88	0.85	0.89	0.84	0.84	0.81	0.88	0.85	0.87	0.85	0.88	0.85	0.87	0.85	0.86	0.85	0.87	0.85
16	V-IND-SG-3-PST - aß	0.85	0.82	0.87	0.86	0.88	0.9	0.87	0.89	0.88	0.9	0.87	0.89	0.86	0.89	0.87	0.89	0.85	0.84	0.88	0.88	0.87	0.88	0.88	0.89	0.87	0.88	0.88	0.88	0.87	0.88
17	V-NFIN - essen	0.85	0.78	0.82	0.79	0.83	0.78	0.82	0.78	0.83	0.78	0.81	0.77	0.78	0.78	0.79	0.77	0.85	0.81	0.83	0.77	0.82	0.77	0.83	0.77	0.82	0.77	0.82	0.77	0.82	0.77
18	V-NFIN-LGSPEC01 - אין	0.72	0.66	0.79	0.76	0.79	0.73	0.79	0.73	0.79	0.73	0.81	0.74	0.76	0.73	0.77	0.74	0.74	0.76	0.8	0.74	0.8	0.74	0.8	0.72	0.81	0.73	0.81	0.73	0.81	0.73
19	V-SBJV-PL-1-PRS - essen	0.86	0.81	0.88	0.85	0.89	0.84	0.88	0.83	0.89	0.84	0.87	0.83	0.85	0.83	0.86	0.83	0.86	0.87	0.89	0.85	0.88	0.84	0.89	0.84	0.88	0.83	0.88	0.84	0.88	0.84
20	V-SBJV-PL-1-PST - äßen	0.85	0.84	0.88	0.87	0.88	0.89	0.88	0.89	0.88	0.9	0.88	0.89	0.86	0.89	0.87	0.89	0.85	0.84	0.88	0.9	0.88	0.89	0.88	0.9	0.88	0.89	0.88	0.89	0.88	0.89
21	V-SBJV-PL-2-PRS - esset	0.86	0.81	0.89	0.87	0.89	0.86	0.89	0.86	0.89	0.86	0.88	0.86	0.86	0.86	0.88	0.86	0.86	0.84	0.89	0.86	0.9	0.86	0.89	0.85	0.9	0.85	0.89	0.86	0.9	0.86
22	V-SBJV-PL-2-PST - äßet	0.82	0.82	0.87	0.85	0.87	0.88	0.87	0.88	0.87	0.88	0.85	0.88	0.85	0.88	0.86	0.88	0.83	0.79	0.87	0.89	0.86	0.89	0.86	0.89	0.87	0.88	0.86	0.89	0.87	0.89
23	V-SBJV-PL-3-PRS - essen	0.87	0.8	0.88	0.86	0.89	0.85	0.88	0.85	0.89	0.85	0.88	0.85	0.86	0.85	0.88	0.85	0.87	0.87	0.89	0.85	0.89	0.85	0.89	0.84	0.89	0.85	0.89	0.85	0.89	0.85
24	V-SBJV-PL-3-PST - äßen	0.86	0.83	0.89	0.87	0.89	0.9	0.89	0.89	0.89	0.9	0.87	0.9	0.87	0.89	0.89	0.9	0.86	0.85	0.89	0.9	0.88	0.9	0.89	0.9	0.88	0.9	0.89	0.9	0.88	0.9
25	V-SBJV-SG-1-PRS - esse	0.86	0.81	0.89	0.86	0.9	0.86	0.89	0.85	0.9	0.86	0.88	0.85	0.85	0.85	0.87	0.85	0.86	0.87	0.9	0.85	0.9	0.85	0.9	0.85	0.9	0.85	0.9	0.85	0.9	0.85
26	V-SBJV-SG-1-PST - äße	0.85	0.82	0.88	0.87	0.89	0.9	0.89	0.89	0.89	0.9	0.88	0.9	0.86	0.89	0.88	0.89	0.85	0.85	0.89	0.9	0.88	0.9	0.89	0.9	0.88	0.9	0.88	0.9	0.88	0.9
27	V-SBJV-SG-2-PRS - essest	0.87	0.79	0.88	0.87	0.88	0.85	0.89	0.85	0.88	0.85	0.88	0.85	0.85	0.85	0.87	0.85	0.87	0.86	0.89	0.85	0.89	0.85	0.89	0.85	0.9	0.85	0.89	0.84	0.89	0.85
28	V-SBJV-SG-2-PST - äßest	0.84	0.84	0.88	0.86	0.88	0.89	0.87	0.88	0.88	0.89	0.87	0.89	0.86	0.88	0.87	0.89	0.85	0.83	0.88	0.89	0.88	0.89	0.88	0.89	0.88	0.89	0.88	0.89	0.88	0.89
29	V-SBJV-SG-3-PRS - esse	0.87	0.79	0.89	0.86	0.88	0.85	0.89	0.84	0.88	0.85	0.88	0.84	0.85	0.84	0.87	0.84	0.87	0.87	0.89	0.84	0.89	0.84	0.89	0.84	0.89	0.84	0.89	0.84	0.89	0.84
30	V-SBJV-SG-3-PST - äße	0.86	0.84	0.88	0.87	0.88	0.9	0.88	0.89	0.88	0.9	0.87	0.89	0.86	0.89	0.88	0.89	0.85	0.85	0.88	0.9	0.88	0.9	0.88	0.9	0.88	0.9	0.88	0.9	0.88	0.9

Figure 6: Reinflection Accuracy scores for the German verb paradigm. Values range from 0.66 to 0.9. Darker teal shades indicate higher accuracy, while darker pink shades reflect lower performance.

	Features	V-IMP-ACT - ät	V-IND-PL-ACT-PRS - äta	V-IND-PL-ACT-PST - åto	V-IND-PL-PASS-PRS - ätas	V-IND-PL-PASS-PST - åtos	V-IND-SG-ACT-PRS - äter	V-IND-SG-ACT-PST - åt	V-IND-SG-PASS-PRS - äts/ätes	V-IND-SG-PASS-PST - åts	V-NFIN-ACT - äta	V-NFIN-PASS - ätas	V-SBJV-ACT-PRS - äte	V-SBJV-ACT-PST - åte	V-SBJV-PASS-PRS - ätes	V-SBJV-PASS-PST - åtes	V-V.CVB-ACT - ätit	V-V.CVB-PASS - ätits	V-V.PTCP-PRS - ätande	V-V.PTCP-PST - äten
1	V-IMP-ACT - ät	0	0.272978304	2.085601578	1.275225012	3.081004263	1.216568047	1.95147929	1.01657982	2.963524396	0.272404264	1.275225012	0.964497041	2.07495069	1.968261487	3.069635244	1.236291913	2.210800568	3.456418384	1.408969805
2	V-IND-PL-ACT-PRS - äta	0.272978304	0	2.087933754	1	3.075793463	1.229495268	2.057176656	1.029843676	2.968735197	0	1	0.952287066	2.077287066	1.957366177	3.064424443	1.27011041	2.206537186	3.196039604	1.461606747
3	V-IND-PL-ACT-PST - åto	2.085601578	2.087933754	0	2.216011369	1.005210801	2.174290221	0.173895899	2.096162956	1.037896731	2.088003157	2.216011369	1.921529968	0.128154574	2.86215064	1.117479867	2.092665615	2.200378967	2.00950495	1.46560142
4	V-IND-PL-PASS-PRS - ätas	1.275225012	1	2.216011369	0	2.087511826	1.263382283	2.187115111	0.274834437	2.058656575	1	0	1.959261014	2.216011369	0.959318827	2.07615894	1.43628612	1.252601703	3.19734974	1.61431316
5	V-IND-PL-PASS-PST - åtos	3.081004263	3.075793463	1.005210801	2.087511826	0	2.932259593	1.150165798	2.069063387	0.145222327	3.075865339	2.087511826	2.905258171	1.116532449	1.906811731	0.112582781	2.986736144	2.078524125	2.844297208	2.323246878
6	V-IND-SG-ACT-PRS - äter	1.216568047	1.229495268	2.174290221	1.263382283	2.932259593	0	2.140378549	1.230222643	2.90052108	1.228887135	1.263382283	1.693611987	2.086750789	1.72382757	2.843202274	1.413643533	2.251539555	3.486732673	1.545051043
7	V-IND-SG-ACT-PST - åt	1.95147929	2.057176656	0.173895899	2.187115111	1.150165798	2.140378549	0	2.059687352	1.005210801	2.057221784	2.187115111	1.890772871	0.173895899	2.833254382	1.150165798	2.107255521	2.213169114	2.011089109	1.485130937
8	V-IND-SG-PASS-PRS - äts/ätes	1.01657982	1.029843676	2.096162956	0.274834437	2.069063387	1.230222643	2.059687352	0	1.958372753	1.029397819	0.274834437	1.70724775	2.091425865	0.952223273	2.052980132	1.258171483	1.209555345	3.454803597	1.426512968
9	V-IND-SG-PASS-PST - åts	2.963524396	2.968735197	1.037896731	2.058656575	0.145222327	2.90052108	1.005210801	1.958372753	0	2.969653864	2.058656575	2.797252487	1.036949313	1.877956481	0.145222327	2.980104216	2.0884579	2.833885471	2.323727185
10	V-NFIN-ACT - äta	0.272404264	0	2.088003157	1	3.075865339	1.228887135	2.057221784	1.029397819	2.969653864	0	1	0.952249408	2.077348066	1.957325747	3.064485538	1.269534333	2.205784732	3.196195006	1.461128387
11	V-NFIN-PASS - ätas	1.275225012	1	2.216011369	0	2.087511826	1.263382283	2.187115111	0.274834437	2.058656575	1	0	1.959261014	2.216011369	0.959318827	2.07615894	1.43628612	1.252601703	3.19734974	1.61431316
12	V-SBJV-ACT-PRS - äte	0.964497041	0.952287066	1.921529968	1.959261014	2.905258171	1.693611987	1.890772871	1.70724775	2.797252487	0.952249408	1.959261014	0	1.820189274	1	2.814779725	1.960962145	2.901942207	3.229306931	2.058144696
13	V-SBJV-ACT-PST - åte	2.07495069	2.077287066	0.128154574	2.216011369	1.116532449	2.086750789	0.173895899	2.091425865	1.036949313	2.077348066	2.216011369	1.820189274	0	2.780672667	1.004263382	2.092665615	2.200378967	1.892277228	1.351531292
14	V-SBJV-PASS-PRS - ätes	1.968261487	1.957366177	2.86215064	0.959318827	1.906811731	1.72382757	2.833254382	0.952223273	1.877956481	1.957325747	0.959318827	1	2.780672667	0	1.814569536	2.129796305	1.94512772	4.153809749	2.231508165
15	V-SBJV-PASS-PST - åtes	3.069635244	3.064424443	1.117479867	2.07615894	0.112582781	2.843202274	1.150165798	2.052980132	0.145222327	3.064485538	2.07615894	2.814779725	1.004263382	1.814569536	0	2.986736144	2.078524125	2.834358732	2.226705091
16	V-V.CVB-ACT - ätit	1.236291913	1.27011041	2.092665615	1.43628612	2.986736144	1.413643533	2.107255521	1.258171483	2.980104216	1.269534333	1.43628612	1.960962145	2.092665615	2.129796305	2.986736144	0	1.001421127	3.582178218	1.349312028
17	V-V.CVB-PASS - ätits	2.210800568	2.206537186	2.200378967	1.252601703	2.078524125	2.251539555	2.213169114	1.209555345	2.0884579	2.205784732	1.252601703	2.901942207	2.200378967	1.94512772	2.078524125	1.001421127	0	3.575958353	2.288184438
18	V-V.PTCP-PRS - ätande	3.456418384	3.196039604	2.00950495	3.19734974	2.844297208	3.486732673	2.011089109	3.454803597	2.833885471	3.196195006	3.19734974	3.229306931	1.892277228	4.153809749	2.834358732	3.582178218	3.575958353	0	2.467584369
19	V-V.PTCP-PST - äten	1.408969805	1.461606747	1.46560142	1.61431316	2.323246878	1.545051043	1.485130937	1.426512968	2.323727185	1.461128387	1.61431316	2.058144696	1.351531292	2.231508165	2.226705091	1.349312028	2.288184438	2.467584369	0

Figure 7: Average edit distances for the Swedish verb paradigm. Values range from 0 to 4.153. Darker red shades indicate closer relationships between feature sets, while darker ball-blue shades represent greater differences.

	Features	V-IMP-ACT - ät	V-IND-PL-ACT-PRS - äta	V-IND-PL-ACT-PST - åto	V-IND-PL-PASS-PRS - ätas	V-IND-PL-PASS-PST - åtos	V-IND-SG-ACT-PRS - äter	V-IND-SG-ACT-PST - åt	V-IND-SG-PASS-PRS - äts/ätes	V-IND-SG-PASS-PST - åts	V-NFIN-ACT - äta	V-NFIN-PASS - ätas	V-SBJV-ACT-PRS - äte	V-SBJV-ACT-PST - åte	V-SBJV-PASS-PRS - ätes	V-SBJV-PASS-PST - åtes	V-V.CVB-ACT - ätit	V-V.CVB-PASS - ätits	V-V.PTCP-PRS - ätande	V-V.PTCP-PST - äten
1	V-IMP-ACT - ät	1	6	52	7	54	6	47	5	48	6	7	6	51	6	53	37	38	66	80
2	V-IND-PL-ACT-PRS - äta	6	1	55	1	51	5	61	7	55	1	1	7	54	6	50	36	34	48	88
3	V-IND-PL-ACT-PST - åto	54	54	1	49	4	56	11	57	13	54	49	57	2	52	5	33	34	116	70
4	V-IND-PL-PASS-PRS - ätas	7	1	51	1	51	6	55	7	55	1	1	7	50	6	50	35	34	33	79
5	V-IND-PL-PASS-PST - åtos	55	50	4	49	1	54	13	54	10	50	49	52	5	52	2	34	31	100	67
6	V-IND-SG-ACT-PRS - äter	6	6	63	6	59	1	65	8	61	6	6	7	53	6	52	39	38	65	93
7	V-IND-SG-ACT-PST - åt	47	55	11	50	13	57	1	50	4	55	50	58	11	53	13	40	41	124	77
8	V-IND-SG-PASS-PRS - äts/ätes	5	8	55	7	53	7	49	1	47	8	7	6	54	5	52	36	36	51	76
9	V-IND-SG-PASS-PST - åts	48	51	13	50	10	56	4	47	1	51	50	53	13	53	10	38	38	103	72
10	V-NFIN-ACT - äta	6	1	55	1	51	5	61	7	55	1	1	7	54	6	50	36	34	48	88
11	V-NFIN-PASS - ätas	7	1	51	1	51	6	55	7	55	1	1	7	50	6	50	35	34	33	79
12	V-SBJV-ACT-PRS - äte	6	7	64	6	57	6	66	5	58	7	6	1	54	1	50	35	32	62	88
13	V-SBJV-ACT-PST - åte	53	53	2	48	5	55	11	56	13	53	48	56	1	51	4	33	34	124	73
14	V-SBJV-PASS-PRS - ätes	7	7	56	6	57	6	57	5	58	7	6	1	49	1	50	34	33	47	79
15	V-SBJV-PASS-PST - åtes	54	49	5	48	2	53	13	53	10	49	48	51	4	51	1	34	31	104	68
16	V-V.CVB-ACT - ätit	38	40	33	36	34	40	39	38	39	40	36	40	33	35	34	1	3	100	60
17	V-V.CVB-PASS - ätits	38	37	36	36	32	40	41	36	37	37	36	39	35	36	32	3	1	88	54
18	V-V.PTCP-PRS - ätande	66	48	137	33	115	68	142	57	119	48	33	64	128	48	108	106	91	1	39
19	V-V.PTCP-PST - äten	81	90	80	79	74	91	85	78	79	90	79	87	73	76	68	56	49	34	1

Figure 8: Edit Script scores for the Swedish verb paradigm. Values range from 1 to 142. Darker purple shades indicate fewer unique character sets (closer relationships), while darker air-force-blue shades reflect greater variation.

	Features	V-IMP-ACT - ät	V-IND-PL-ACT-PRS - äta	V-IND-PL-ACT-PST - åto	V-IND-PL-PASS-PRS - ätas	V-IND-PL-PASS-PST - åtos	V-IND-SG-ACT-PRS - äter	V-IND-SG-ACT-PST - åt	V-IND-SG-PASS-PRS - äts/ätes	V-IND-SG-PASS-PST - åts	V-NFIN-ACT - äta	V-NFIN-PASS - ätas	V-SBJV-ACT-PRS - äte	V-SBJV-ACT-PST - åte	V-SBJV-PASS-PRS - ätes	V-SBJV-PASS-PST - åtes	V-V.CVB-ACT - ätit	V-V.CVB-PASS - ätits	V-V.PTCP-PRS - ätande	V-V.PTCP-PST - äten
1	V-IMP-ACT - ät	0.88	0.87	0.82	0.86	0.81	0.87	0.82	0.87	0.82	0.86	0.87	0.85	0.82	0.85	0.82	0.82	0.82	0.85	0.81
2	V-IND-PL-ACT-PRS - äta	0.73	0.83	0.7	0.82	0.69	0.73	0.7	0.71	0.69	0.83	0.82	0.81	0.7	0.82	0.69	0.69	0.68	0.81	0.69
3	V-IND-PL-ACT-PST - åto	0.75	0.74	0.84	0.74	0.83	0.75	0.83	0.74	0.82	0.75	0.75	0.75	0.84	0.73	0.83	0.8	0.8	0.77	0.8
4	V-IND-PL-PASS-PRS - ätas	0.72	0.82	0.69	0.83	0.68	0.71	0.69	0.7	0.69	0.82	0.83	0.81	0.69	0.81	0.69	0.7	0.7	0.82	0.7
5	V-IND-PL-PASS-PST - åtos	0.73	0.73	0.78	0.72	0.8	0.73	0.79	0.73	0.79	0.72	0.73	0.71	0.79	0.71	0.8	0.78	0.77	0.74	0.78
6	V-IND-SG-ACT-PRS - äter	0.82	0.81	0.77	0.81	0.75	0.83	0.78	0.79	0.76	0.81	0.81	0.82	0.76	0.81	0.75	0.77	0.77	0.8	0.75
7	V-IND-SG-ACT-PST - åt	0.76	0.75	0.81	0.74	0.79	0.75	0.82	0.75	0.8	0.75	0.75	0.75	0.81	0.74	0.8	0.77	0.77	0.75	0.79
8	V-IND-SG-PASS-PRS - äts/ätes	0.8	0.8	0.76	0.8	0.76	0.79	0.76	0.8	0.75	0.81	0.8	0.78	0.74	0.77	0.75	0.75	0.75	0.77	0.73
9	V-IND-SG-PASS-PST - åts	0.72	0.7	0.78	0.71	0.78	0.71	0.79	0.7	0.79	0.7	0.71	0.71	0.79	0.71	0.78	0.75	0.76	0.72	0.75
10	V-NFIN-ACT - äta	0.72	0.84	0.7	0.82	0.69	0.72	0.7	0.71	0.7	0.84	0.82	0.82	0.7	0.81	0.7	0.7	0.7	0.81	0.7
11	V-NFIN-PASS - ätas	0.69	0.79	0.66	0.78	0.65	0.68	0.66	0.66	0.66	0.79	0.8	0.76	0.67	0.77	0.65	0.69	0.68	0.76	0.68
12	V-SBJV-ACT-PRS - äte	0.71	0.8	0.68	0.79	0.68	0.71	0.68	0.69	0.68	0.8	0.8	0.81	0.69	0.8	0.68	0.7	0.69	0.78	0.67
13	V-SBJV-ACT-PST - åte	0.77	0.77	0.83	0.76	0.83	0.76	0.82	0.75	0.81	0.77	0.76	0.74	0.84	0.74	0.83	0.8	0.79	0.76	0.81
14	V-SBJV-PASS-PRS - ätes	0.69	0.77	0.68	0.77	0.68	0.69	0.67	0.68	0.67	0.77	0.77	0.77	0.67	0.77	0.66	0.67	0.67	0.75	0.68
15	V-SBJV-PASS-PST - åtes	0.67	0.66	0.74	0.66	0.73	0.65	0.74	0.67	0.74	0.68	0.67	0.66	0.75	0.66	0.75	0.69	0.69	0.66	0.71
16	V-V.CVB-ACT - ätit	0.76	0.76	0.8	0.76	0.79	0.77	0.8	0.76	0.79	0.76	0.76	0.76	0.8	0.75	0.79	0.82	0.81	0.77	0.8
17	V-V.CVB-PASS - ätits	0.71	0.73	0.74	0.72	0.75	0.71	0.75	0.73	0.74	0.72	0.73	0.7	0.73	0.71	0.75	0.78	0.77	0.71	0.73
18	V-V.PTCP-PRS - ätande	0.65	0.74	0.64	0.74	0.63	0.65	0.64	0.62	0.63	0.75	0.74	0.73	0.65	0.72	0.63	0.64	0.63	0.8	0.68
19	V-V.PTCP-PST - äten	0.71	0.7	0.74	0.69	0.73	0.69	0.75	0.7	0.73	0.7	0.69	0.7	0.74	0.69	0.74	0.74	0.73	0.75	0.8

Figure 9: Reinflection Accuracy scores for the Swedish verb paradigm. Values range from 0.62 to 0.88. Darker teal shades indicate higher accuracy, while darker pink shades reflect lower performance.