Close, but different: A data perspective on aspect in Russian and Ukrainian

Yulia Zinova
Heinrich Heine University
zinova@hhu.de

Inna Stupak
University of Tübingen
inna.stupak@uni-tuebingen.de

Heinrich Heine Universität Düsseldorf

Eberhard Karls Universität Tübingen

Overview

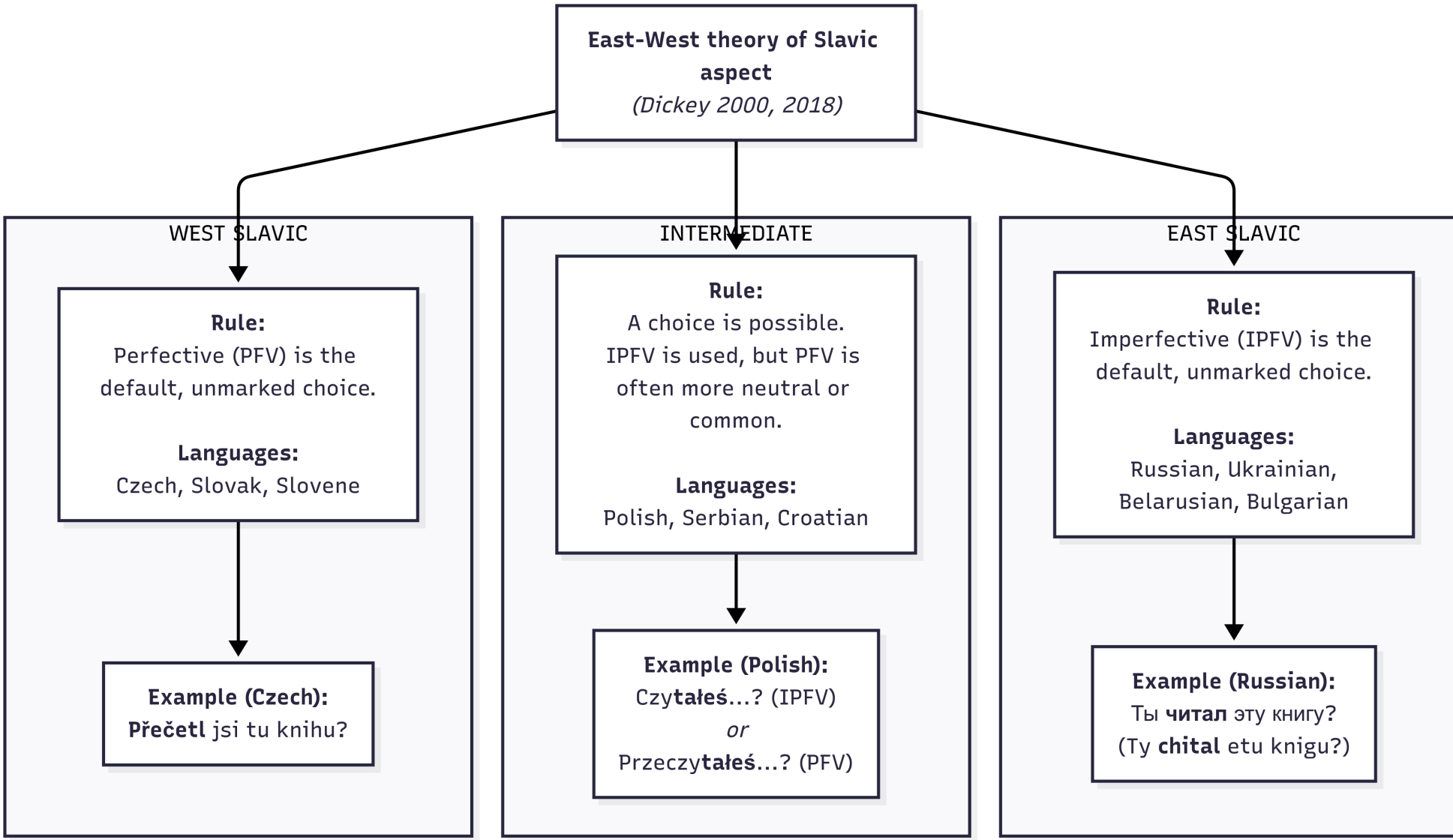
1. Theory of aspect
   - Slavic languages vs. English/other languages
   - Variation within Slavic languages
   - Lexical vs. Grammatical aspect
2. Experiment 1: All verbs
   - Method
   - Results
3. Experiment 2: Aligned verbal pairs
   - Data
   - Method
   - Results
   - Interpretation
4. Experiment 3: Difference vectors
   - Visualization
   - Statistical analysis
5. Discussion

General Approaches to Aspect

- Slavic languages are considered together in opposition with other languages.
- Slavic languages typically distinguish between perfective and imperfective aspects.
- The role of aspect in Slavic languages is different from that in English.

Feature	**Slavic Languages**	**English**
Strong aspectual opposition	Yes	No
Perfective–Imperfective	Core distinction	Often auxiliary-based
Role of Aspect	Morphologically central	More peripheral

East-West theory of Slavic aspect



Lexical vs. Grammatical Aspect

Grammatical Aspect	**Lexical Aspect**
• Perfective = completed • Imperfective = ongoing • Expressed via (inflectional) morphology	• States • Activities • Accomplishments • Achievements • Based on Vendler & Smith

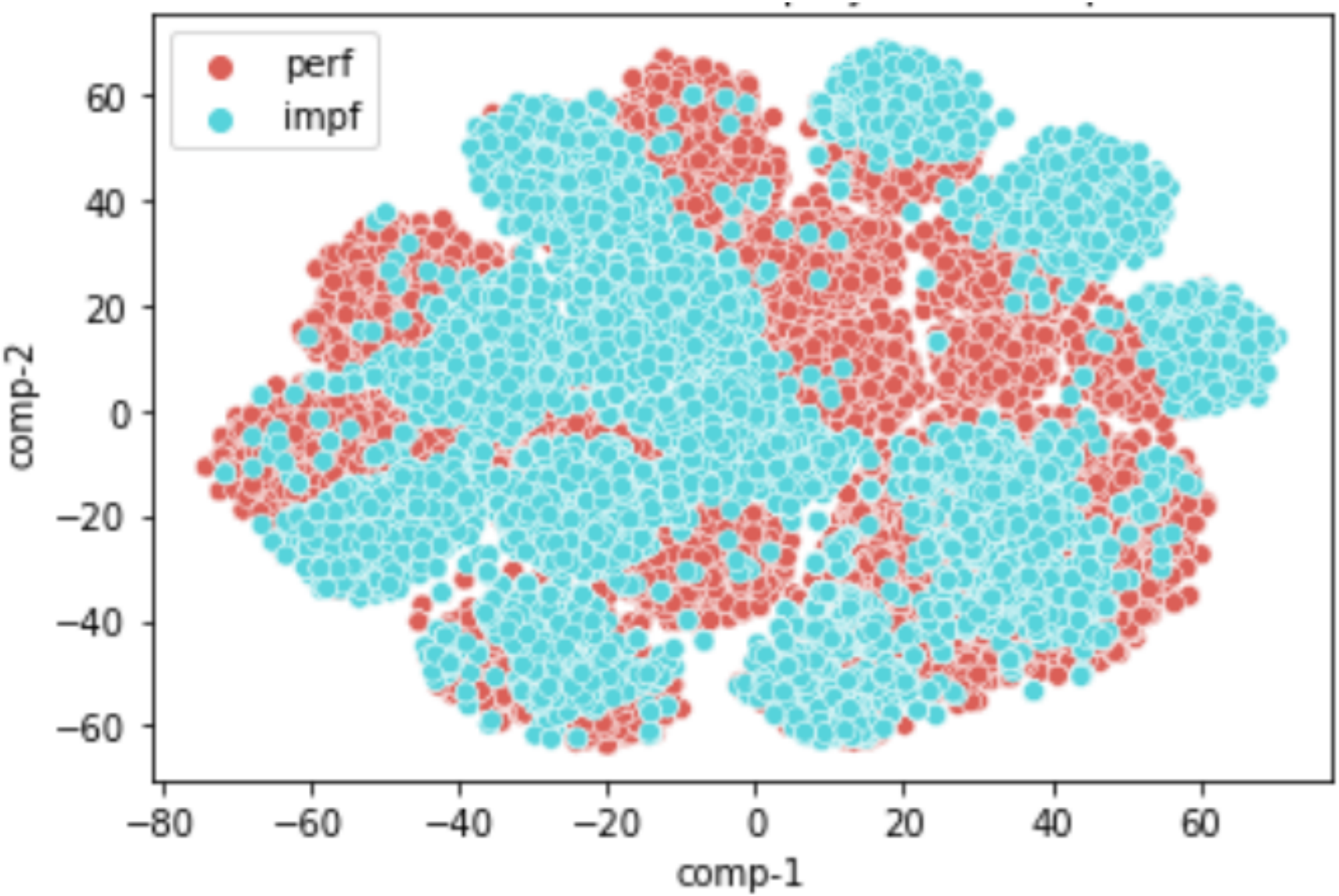
Aspect in Russian and Ukrainian

- **Key question:** Are Russian and Ukrainian aspect systems structurally different?
- Some studies highlight similarities between Russian and Ukrainian.
- Influence from Polish may introduce differences in Ukrainian.
- This study uses a data-driven approach to explore this.

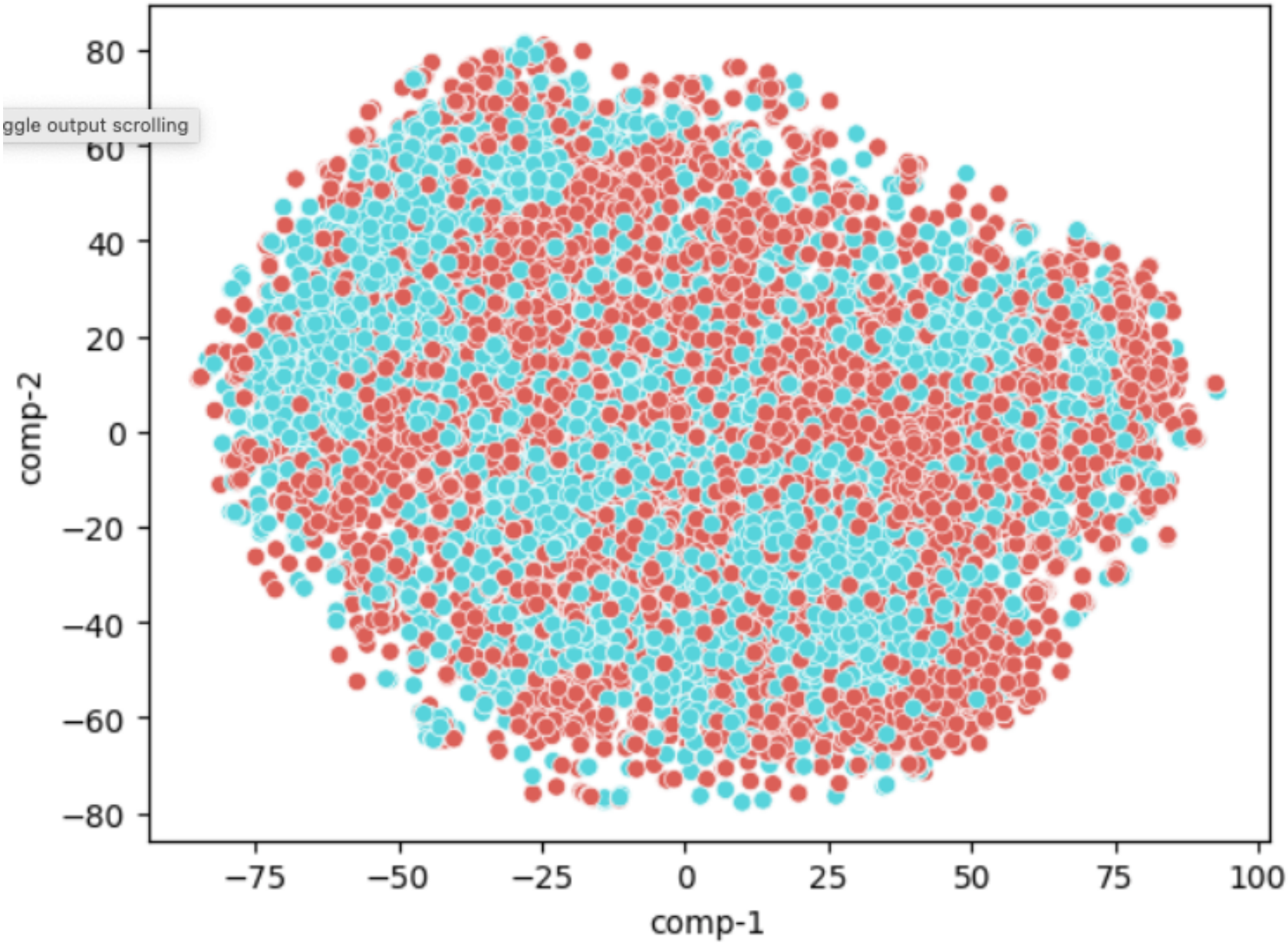
Pipeline

1. Extract Russian and Ukrainian verbs from dictionary data ($>$ 1000 verbs)
2. Generate a set of (personal) forms using the Morphological Analyzer and Generator for Russian and Ukrainian Languages (pymorphy2, Korobov 2015)
3. Extract fastText vectors (Bojanowski et al., 2017) for each form
4. Perform dimensionality reduction with PCA + t-SNE
5. Visualize and color by aspect: red for perfective and blue for imperfective

Visualization of Russian Verbs



Visualization of Ukrainian Verbs



Data

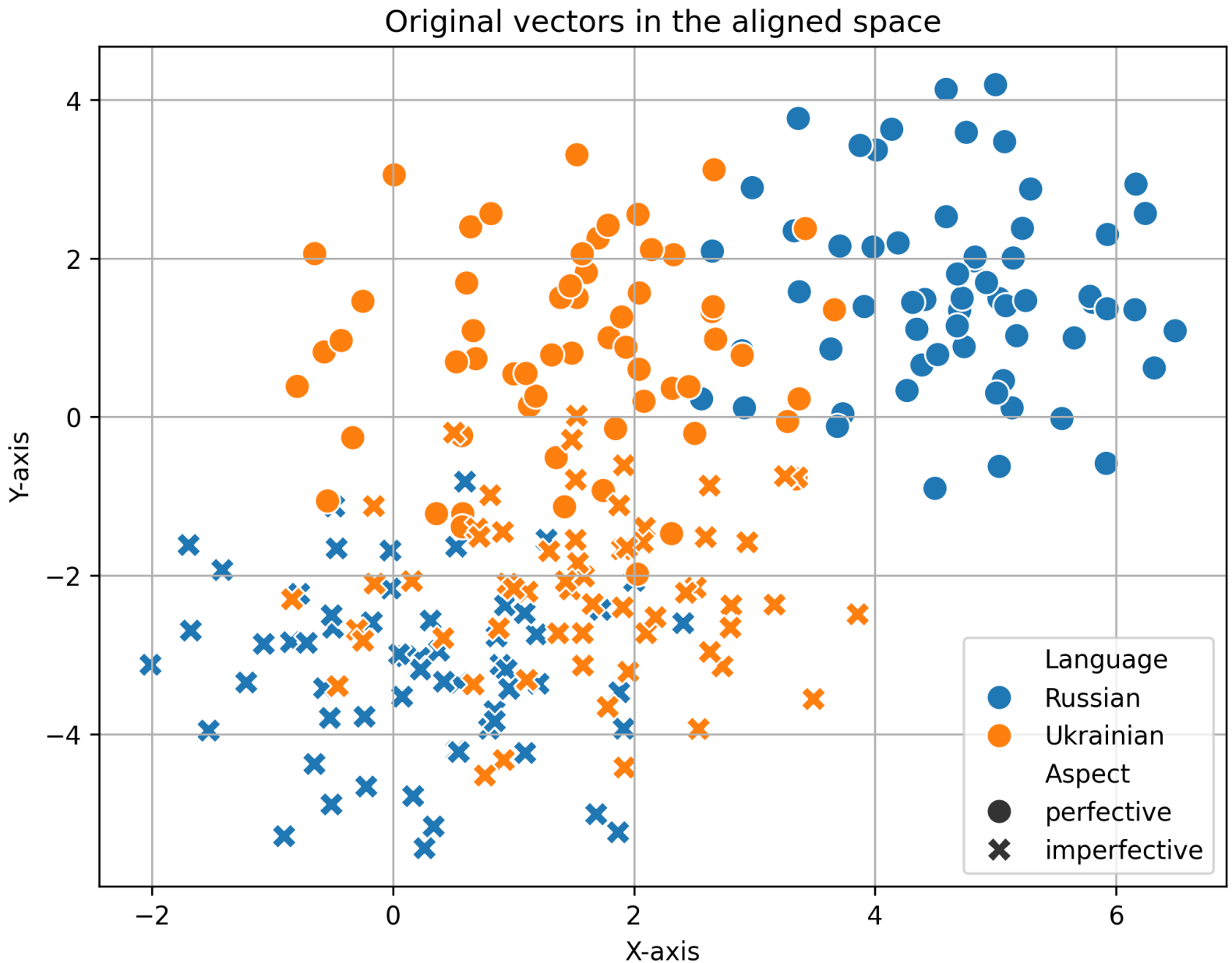
- Data collected from Russian and Ukrainian corpora.
- Focus on aligned verbal pairs with aspectual information.
- Manual annotation used to ensure quality.

Russian		Ukrainian		English
imperfective	perfective	imperfective	perfective	
pisat'	*napisat'*	*pysaty*	*napysaty*	to write
rešat'	*rešit'*	*vyrišuvaty*	*vyryšyty*	to decide
govorit'	*pogovorit'*	*hovoryty*	*pohovoryty*	to talk
brat'	*vzyat'*	*braty*	*vzyaty*	to take
pisat'	*popisyvat'*	*pysaty*	*popysuvaty*	to write
zažech'	*zažigat'*	*zapalyty*	*zapalyuvaty*	to lit

Visualization and statistical analysis for aligned verbal pairs

- Goal: Test whether Russian and Ukrainian encode aspect differently in embedding space.
- Selected 100 Ukrainian aspectual verb pairs; translated to Russian.
- Filtered to keep only pairs with:
  - Trained vectors for all four forms (perf/imperfective in both languages).
  - Distinct translations.
- Result: 62 aligned quadruples (RU+UKR, perf+imperf).
- Aligned RU and UKR vector spaces using Smith et al. 2017.
- Validation:
  - Precision: 0.84
  - Cosine similarity: 0.63

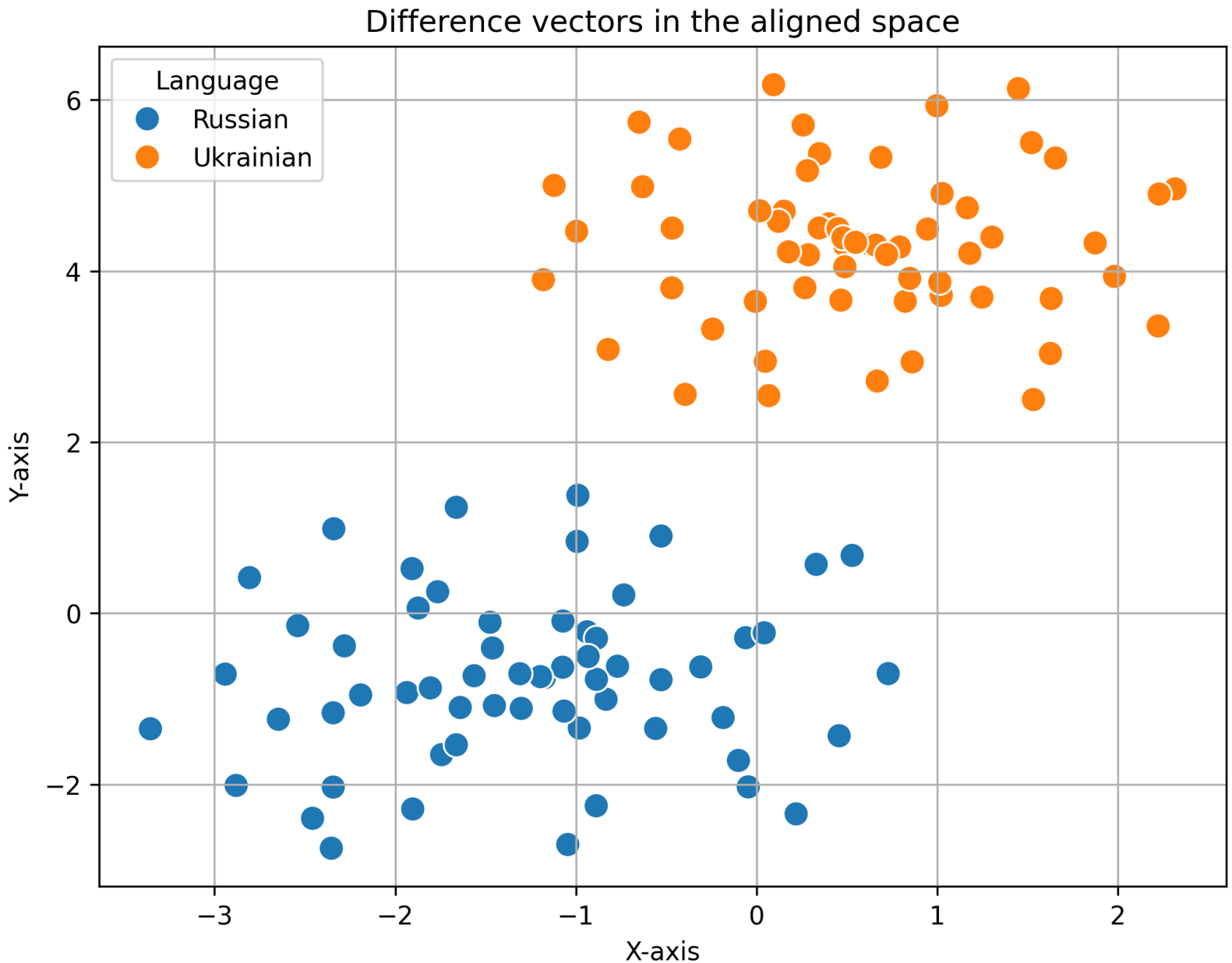
Embedding Space Visualization



Embedding Space Visualization Interpretation

- Russian and Ukrainian verbs overlap but show distinct patterns.
- Ukrainian (orange): Spread across center and right.
- Russian (blue): Cluster in upper-right and lower-left regions.
- Indicates persistent language-specific features even after alignment.
- Closer placement of Ukrainian perf/imperf forms suggests:
  - Stronger semantic or morphological similarity.
  - Possibly more consistent derivation.
- Russian pairs show more separation $\rightarrow$ more distinct aspectual encoding.

Difference Vectors Visualization



Difference Vectors Statistical Test

- Computed difference vectors (perf − imperf) for each pair.
- The visualization shows a clear separation between Russian and Ukrainian.
- Applied **Wasserstein Distance Test**:
  - p-value $\ll 0.001 \rightarrow$ distributions significantly different.
- Confirms that aspectual shifts behave differently in Russian and Ukrainian.
- Interpretation:
  - Ukrainian: more systematic aspect derivation.
  - Russian: greater morphological and semantic diversity in aspect.

Gradient Nature of Aspectual Distinction

- Adopt a **gradient view** of aspect: not strictly grammatical vs. lexical.
- Embedding patterns reflect degree of grammaticalization:
  - **Russian**: clearer separation → stronger grammatical marking.
  - **Ukrainian**: more overlap → context-sensitive, semantically flexible.
- Supports continuum view: aspect interacts with telicity, boundedness, habituality (Barentsen et al., 2015; Dickey & Kresin, 2009).
- Embeddings capture this difference without enforcing hard categories.

Interpretation and Theoretical Implications

- Wasserstein distance test confirms language-specific distributions of aspect pairs.
- Aspectual encoding reflects:
  - **Russian**: Predominantly grammatical system (Pereltsvaig, 2008).
  - **Ukrainian**: More lexical, similar to Polish (Divjak et al., 2024).
- Aligns with previous typological findings (Ghorodensjka, 2019).
- Embedding behavior mirrors deeper syntactic, semantic, and lexical differences.

Barentsen, A., R. Genis, M. van Duijkeren-Hrabova, J. Kalsbeek, R. Lučić et al. 2015. V poiskakh skhodstv i razlichiĭ mezhdu russkim, pol'skim, cheshkim i serbochorvatskim iazykami pri vybore vida v sluchaiakh ogranichennoĭ kratnosti. *Die Welt der Slaven. Sammelbände* 56.

Bojanowski, P., E. Grave, A. Joulin & T. Mikolov. 2017. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5. 135–146.

Dickey, S. M. & S. C. Kresin. 2009. Verbal aspect and negation in Russian and Czech. *Russian linguistics* 33(2). 121–176.

Divjak, D., I. Testini & P. Milin. 2024. On the nature and organisation of morphological categories: verbal aspect through the lens of associative learning. *Morphology* 34. 243–280. doi:10.1007/s11525-024-09423-0.

Ghorodensjka, K. 2019. Morfologhichni kateghoriji ukrajinsjkoji movy v novykh semantyko-ghramatychnykh vymirakh .

Korobov, M. 2015. Morphological Analyzer and Generator for Russian and Ukrainian Languages, doi:10.1007/978-3-319-26123-2_31.

Pereltsvaig, A. 2008. Aspect in Russian as grammatical rather than lexical notion: Evidence from Heritage Russian. *Russian linguistics* 32(1). 27–42.

Smith, S. L., D. H. P. Turban, S. Hamblin & N. Y. Hammerla. 2017. Offline bilingual word vectors, orthogonal transformations and the inverted softmax. In

International Conference on Learning Representations, https://openreview.net/forum?id=r1Aab85gg.