What about Morphology?

- Up to now, we have focussed on dependency and part-of-speech information and neglected inflectional morphology.
- When dealing with languages like Chinese and English, this seems reasonable.
- But what about languages with rich inflectional morphology? Does parsing benefit from the presence of morphological information in word representations?

Estonian: Inflectional paradigm of *kala* 'fish' with 29 members; data due to Loo (2017)

Case	Singular	Plural	English translation
Nominative	kala	kalad	fish (subject)
Genitive	kala	kalade	of a fishfish (total object)
Partitive	kala	kalasid, kalu	fish (partial object)
Illative	kalasse	kaladesse	into a fish
Inessive	kalas	kalades	in a fish
Elative	kalast	kaladest	from a fish
Allative	kalale	kaladele	onto a fish
Adessive	kalal	kaladel	on a fish
Ablative	kalalt	kaladelt	from a fish
Translative	kalaks	kaladeks	[to turn] into a fish
Terminative	kalani	kaladeni	up to a fish
Essive	kalana	kaladena	as a fish
Abessive	kalata	kaladeta	without a fish
Comitative	kalaga	kaladega	with a fish

Estonian: paradigm of *jalg* 'foot' with 46 members; data due to Loo (2017)

Case	Sing.	Plural	English translation
Nominative	**jalg**	**jalad**	foot (subject)
Genitive	**jala**	**jalgade**, **jalge**	of a foot (total object)
Partitive	**jalga**	**jalgasid, jalgu**	foot (as a partial object)
Illative-1	**jalga**	-	into a foot
Illative-2	jalasse	**jalgadesse**, jalusse, jalgesse	into a foot
Inessive	**jalas**	**jalgades**, **jalus**, **jalges**	in a foot
Elative	**jalast**	**jalgadest**, **jalust**, **jalgest**	from a foot
Allative	**jalale**	**jalgadele**, **jalule**, **jalgele**	onto a foot
Adessive	**jalal**	**jalgadel**, **jalul**, **jalgel**	on a foot
Ablative	**jalalt**	**jalgadelt**, jalult, **jalgelt**	from a foot
Translative	**jalaks**	**jalgadeks**, jaluks, jalgeks	[to turn] into a foot
Terminative	jalani	**jalgadeni**, jalgeni	up to a foot
Essive	jalana	jalgadena	as a foot
Abessive	**jalata**	**jalgadeta**	without a foot
Comitative	**jalaga**	**algadega**	with a foot

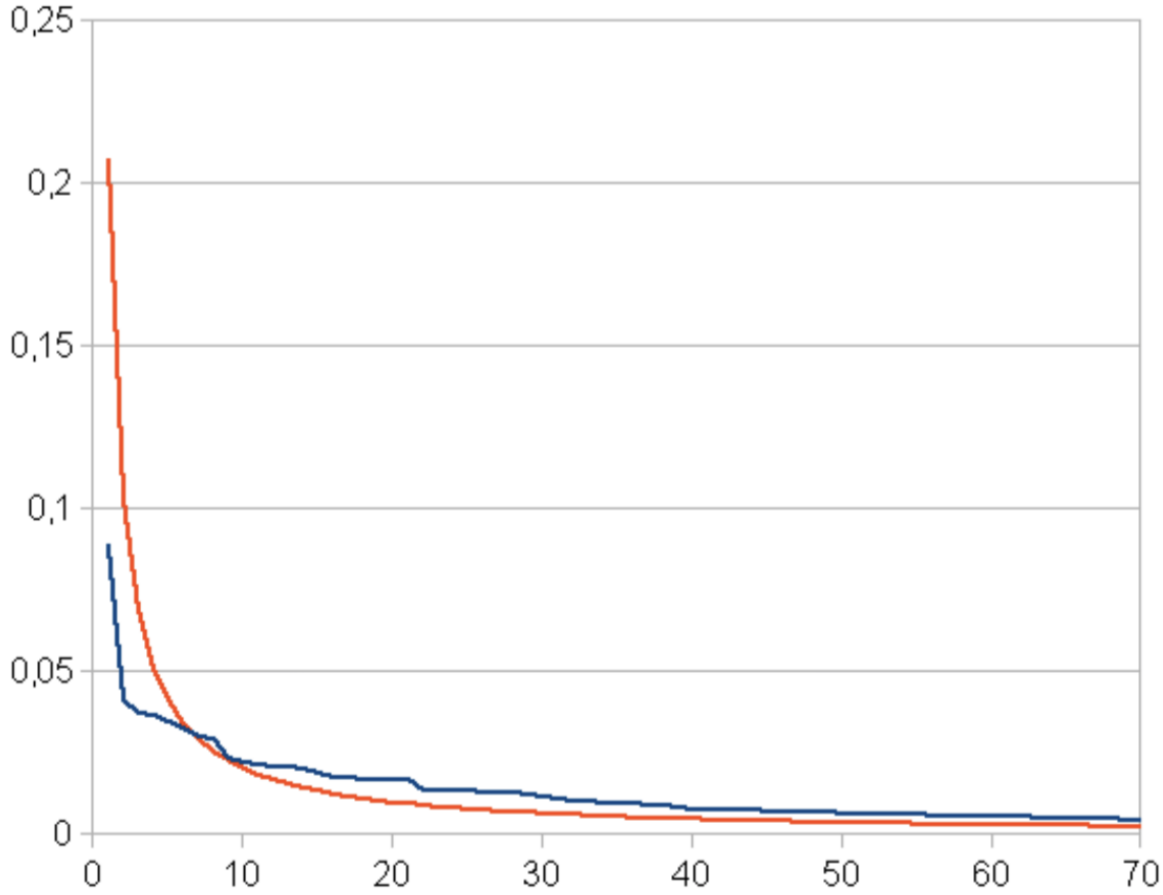
forms attested in a balanced corpus of Estonian rendered in bold

Distribution of Case Inflections in Russian

	No. of types	No. of tokens	Nom	Gen	Dat	Acc	Ins	Loc
1. All nouns	9.073*	102.173	33.6%	24.6%	5.1%	19.5%	7.8%	9.4%
2. Common nouns	9.073	89.384	28.3	26.0	5.0	21.8	8.6	10.3
3. Proper nouns	?	12.789	76.3	13.5	5.5	1.1	1.4	2.2
4. Personal common individual	119	11.769	54.1	22.5	6.9	7.0	8.0	1.5
5. Personal collective	29	2.565	23.9	48.0	4.2	9.6	6.2	8.3
6. Animal	9	746	35.6	28.4	3.8	21.6	6.0	4.6
7. Body parts	31	3.318	18.2	9.9	3.2	36.5	20.3	11.9
8. Concrete count	116	5.475	23.0	20.7	4.3	32.0	9.4	10.5
9. Concrete mass	25	1.565	21.3	31.6	2.2	24.3	13.6	6.9
10. Non-enduring objects	31	2.127	34.5	19.0	4.1	21.5	10.8	8.8
11. Abstract qualities	21	1.295	33.3	24.9	3.8	17.4	12.3	9.0
12. Place nouns	87	7.747	11.8	30.6	6.0	24.4	3.3	23.8
13. Place institutions	17	2.445	13.0	40.9	2.3	17.8	1.8	24.1
14. Time periods	8	2.998	12.8	37.5	2.0	36.0	3.4	8.3
15. Measures	7	480	2.7	85.4	0.8	5.2	1.2	4.6
16. First and second person pronouns	4	15.901	51.8	21.8*	14.5	9.5	2.2	0.2

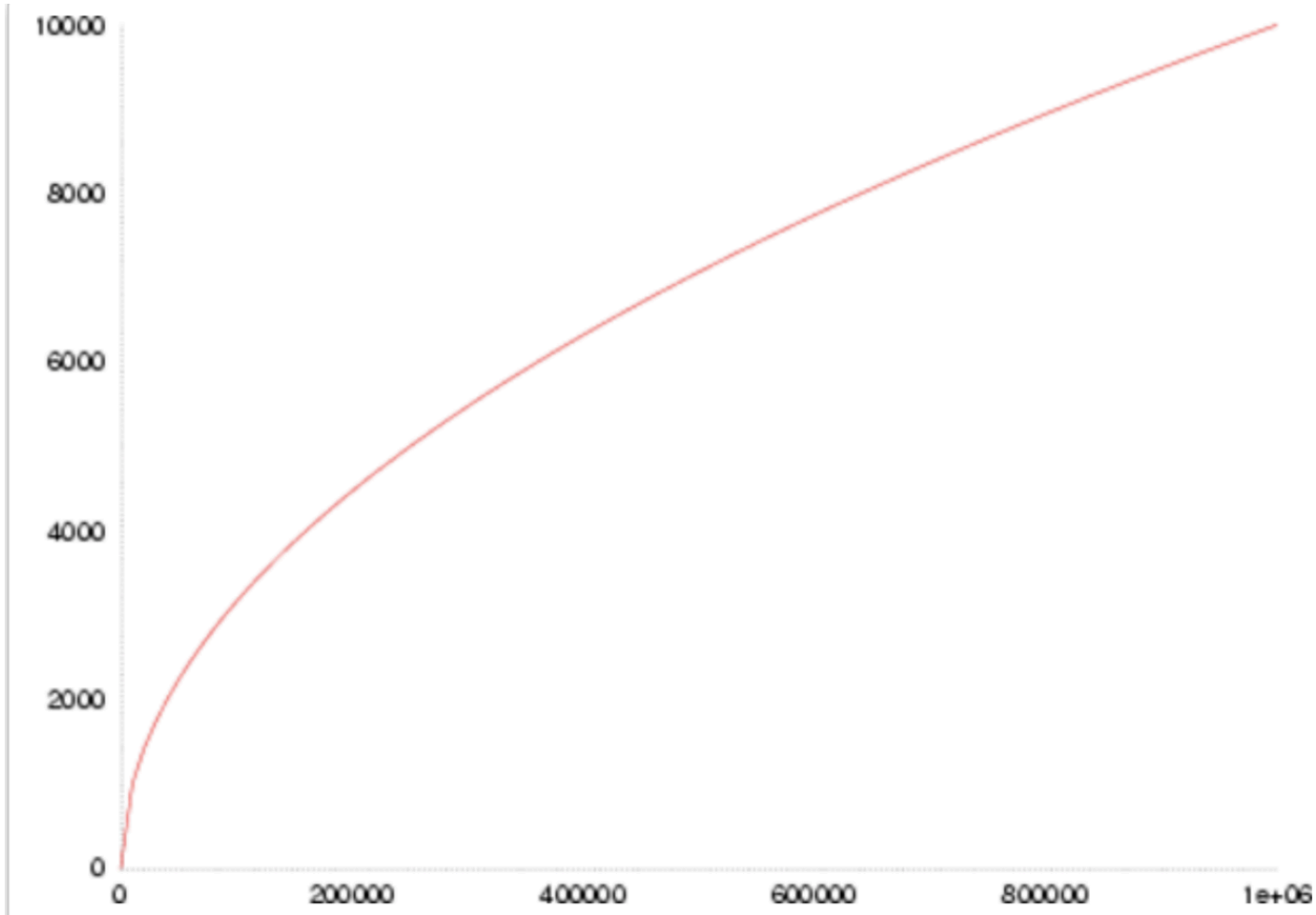
Table 3. The relation of Russian case frequencies to semantic features, adopted from Greenberg (1974,1991: 210).

Zipf's Law: Word Frequency Distributions



$cf_i \propto \frac{1}{i}$, where cf_i is the ith-most common term

Heap's Law: Vocabulary size as a function of number of tokens



$$V = KT^B$$

Paradigm Frequency Distributions

Corpus	Tokens (millions)	Infl. categories	Max. infl. categories per lemma	Max. saturation
Basque	0.6	22	16	72.7
Catalan	1.7	45	33	73.3
Czech	2.0	72	41	56.9
English (Brown Corpus)	1.2	6	6	100.0
English (Wall Street Journal Corpus)	1.3	6	6	100.0
Finnish	2.1	365	147	40.3
Greek	2.8	83	45	54.2
Hebrew	2.5	33	23	69.7
Hungarian	1.2	76	48	63.2
Italian	1.4	55	47	85.5
Slovene	2.4	32	24	75.0
Spanish	2.6	51	34	66.7
Swedish	1.0	21	14	66.7

This table is taken from Yang (2015). *The Price of Linguistic Productivity*. MIT Press.

Paradigm Frequency Distributions – Childes

Corpus	Tokens (millions)	Infl. categories	Max. infl. categories per lemma	Max. saturation
CHILDES Catalan	0.3	39	27	69.2
CHILDES Italian	0.3	49	31	63.3
CHILDES Spanish	1.4	55	46	83.6

Paradigm Frequency Distributions

	COSMAS II	DECOW16
{ich,{er,sie,es}} verlief	139.487	136.055
du verliefst	0	1
{wir, sie} verliefen	27.802	26.841
ihr verlieft	20	105

Conclusion: Even in very large corpora, coverage of morphological paradigms will be incomplete.

Morphological Analysis – Computational Linguistics Methods

- Finite-State Morphology
  - uses finite-state transducers for analyzing and generating inflected forms item advantage: theoretically sound, high-precision output, computationally efficient
  - however: requires hand-crafted rules and may not be available for many low(er)-resourced languages.
- Stemming
  - uses heuristic rules to strip productive prefixes and/or suffixes from inflected word forms
  - advantage: can be implemented with comparatively low human effort
  - however: noisy output, especially for irregular forms or for patterns with low productivity

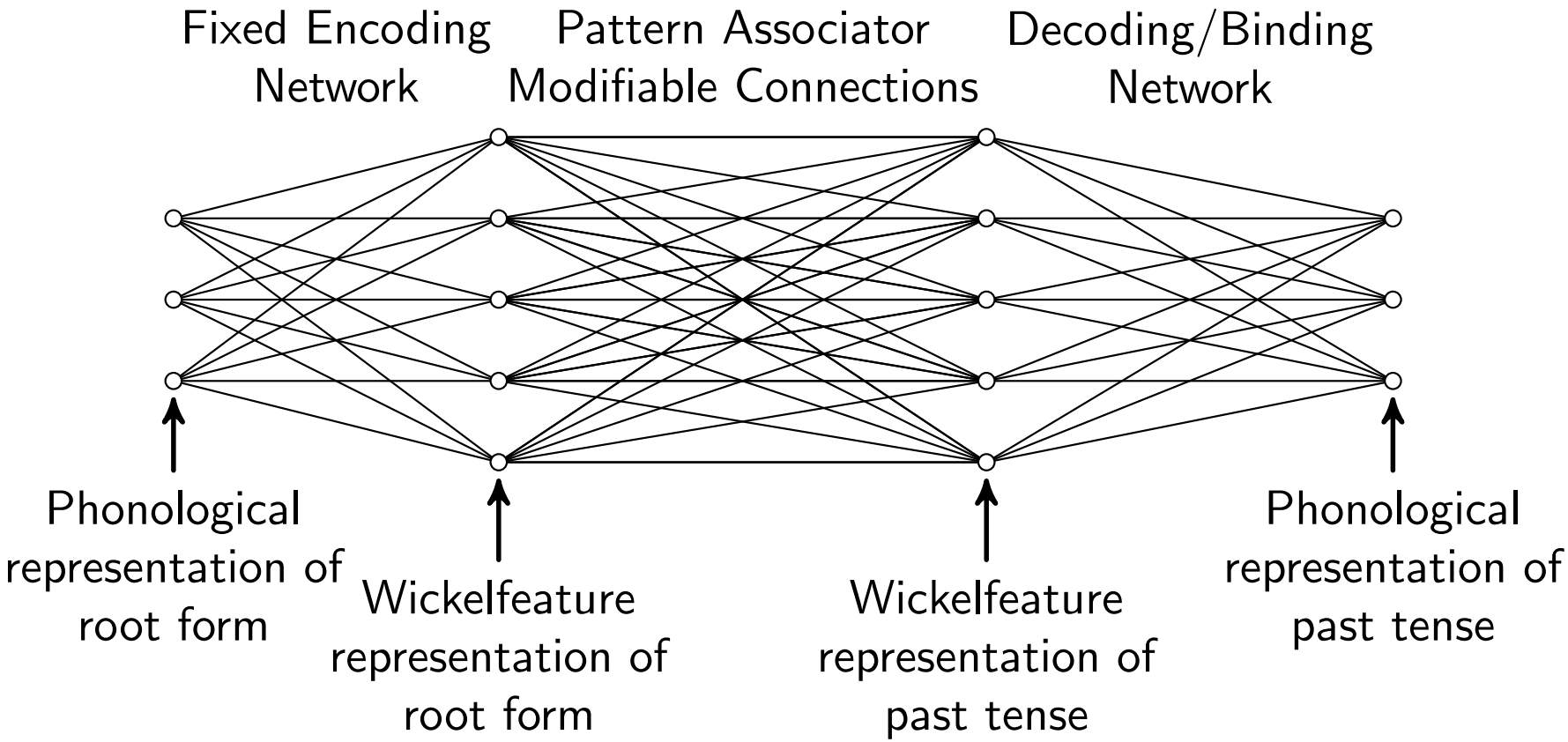
Neural Approaches to Learning Morphology

Using neural networks (simple feed-forward networks) to learn morphological patterns: the past tense in English (Rumelhart and McClelland (1986)

- requires training data of present tense-past tense pairs as input-output pairs for training the network.
- phoneme sequences for present tense and past tense tense forms are encoded as *Wickelfeatures*, whose associations are learnt by the network.

Neural approaches ("deep learning") approaches have made a big comeback in machine learning in general and in computational linguistics in particular.

Neural Approaches to Learning Morphology



Children Learning Morphology

THIS IS A WUG.


NOW THERE IS ANOTHER ONE.

THERE ARE TWO OF THEM.

THERE ARE TWO _____.

From: J. Berko (1958) 'The Childs Learning of English Morphology', Word 14: 150-177.

Neural Approaches to Learning Morphology

References:
Rumelhart, D.E. and McClelland, J.L. (1986). On learning the past tenses of English verbs. In McClelland, J.L. et al., eds. Parallel Distributed Processing: Explorations in the Microstructures of Cognition (Vol. 2),. MIT Press, pp. 216271,
Pinker, S. and Prince, A. (1988) On language and connectionism: analysis of a parallel distributed processing model of language acquisition. Cognition 28, 73193