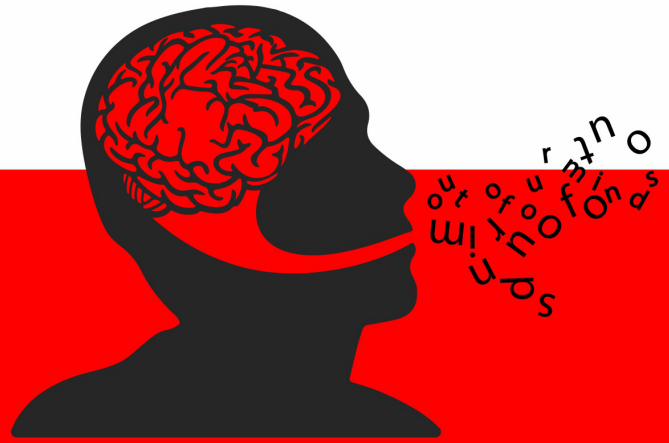
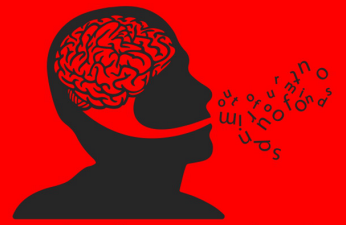
Behavioural Profiles

Dagmar Divjak
<d.divjak@bham.ac.uk>

UNIVERSITY OF BIRMINGHAM

LEVERHULME TRUST





Behavioral Profiling

“the recording and analysis of a word's behavioural characteristics, so as to assess or predict their capabilities in a certain sphere or to assist in identifying categories of words

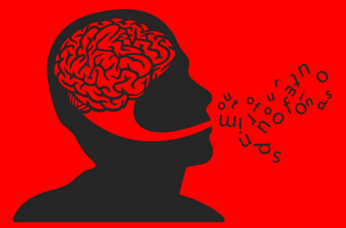
out of our minds

Behavioral Profiling

= “find the man who wasn’t there”

= find the word that could be used

= chart behaviour of *word* across contexts on a *multitude* of dimensions

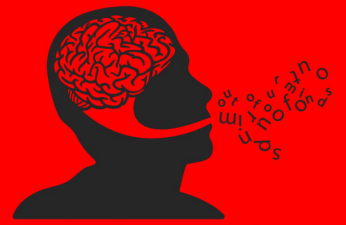


Distributional hypothesis

- Harris 1954
  - meaning is a function of distribution
  - the meaning of a word derives (probabilistically) from the linguistic contexts in which it occurs

Context reveals meaning

Similarity in context implies similarity in meaning

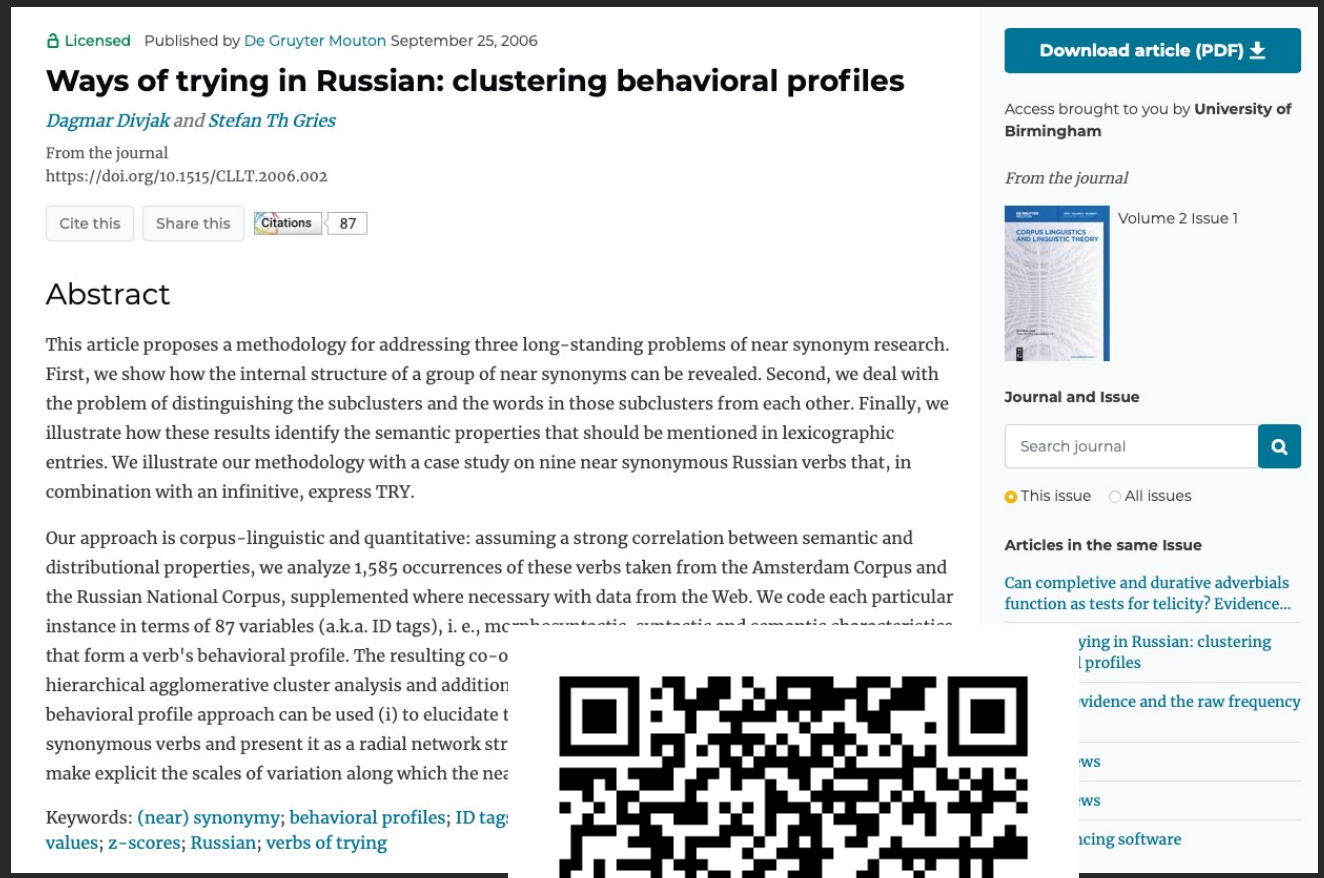


Similar ideas

- Firth (1957: 11)
  - “You shall know a word by the company it keeps.”
- Bolinger (1968: 127)
  - “A difference in syntactic form always spells a difference in meanings.”
- Cruse (1986: 1)
  - “The semantic properties of a lexical item are fully reflected in appropriate aspects of the relations it contracts with actual and potential contexts.”

out of our minds

BPs



Licensed Published by De Gruyter Mouton September 25, 2006

Ways of trying in Russian: clustering behavioral profiles

Dagmar Divjak and *Stefan Th Gries*

From the journal
https://doi.org/10.1515/CLLT.2006.002

Cite this | Share this | Citations 87

Abstract

This article proposes a methodology for addressing three long-standing problems of near synonym research. First, we show how the internal structure of a group of near synonyms can be revealed. Second, we deal with the problem of distinguishing the subclusters and the words in those subclusters from each other. Finally, we illustrate how these results identify the semantic properties that should be mentioned in lexicographic entries. We illustrate our methodology with a case study on nine near synonymous Russian verbs that, in combination with an infinitive, express TRY.

Our approach is corpus-linguistic and quantitative: assuming a strong correlation between semantic and distributional properties, we analyze 1,585 occurrences of these verbs taken from the Amsterdam Corpus and the Russian National Corpus, supplemented where necessary with data from the Web. We code each particular instance in terms of 87 variables (a.k.a. ID tags), i. e., mo
that form a verb's behavioral profile. The resulting co-o
hierarchical agglomerative cluster analysis and addition
behavioral profile approach can be used (i) to elucidate t
synonymous verbs and present it as a radial network str
make explicit the scales of variation along which the nea

Keywords: (near) synonymy; behavioral profiles; ID tag
values; z-scores; Russian; verbs of trying

Download article (PDF)

Access brought to you by **University of Birmingham**

From the journal

Volume 2 Issue 1

Journal and Issue

Search journal

This issue | All issues

Articles in the same Issue

Can completive and durative adverbials function as tests for telicity? Evidence...

ying in Russian: clustering
l profiles

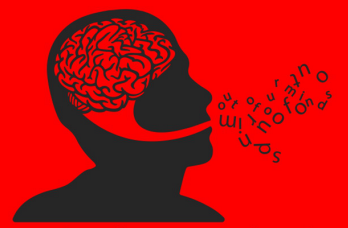
vidence and the raw frequency

ws

ws

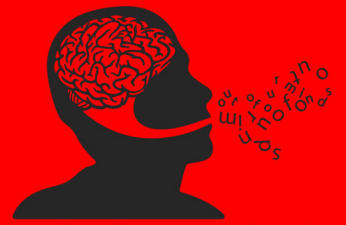
cing software

Divjak (2004)
Divjak & Gries (2006)
https://www.academia.edu/12926961/Ways_of_trying_in_Russian_clustering_behavioral_profiles



Prerequisites

- Allows you to capture *usage*
  - Detailed snapshot
- Research question needs to be
  - couched in usage-based theories
    - Bottom-up, data-driven
  - phrased in terms of *usage*, not “feature” analysis
    - NOT few hand-picked pre-determined features



Application range of BPs

- Method successfully applied to *determine core of* and *distinguish* between
  - polysemous, synonymous and antonymous
  - prefixes, adjectives, nouns, verbs, adverbs and constructions
  - in L1 and L2
  - within and across languages

(McEnery & Hardie 2012; Lehecka 2015; Bębeniec 2024)

Don’t You Try ...

Но **Сирота** все еще **силился** **что-то сказать**, и снова невозможно было понять ни слова из того, что он говорил. Малинин наконец не выдержал и прекратил эту обоюдную муку: **Ты не** **старайся**, Сирота, все равно я не понимаю: у тебя рот разбитый ... звук и только, а голоса нет. В госпитале полежишь - восстановится, а сейчас **не** **пробуй**, не мучь себя (...) [К. Симонов. Живые и мертвые]

But Sirota was still **trying/making efforts** to say something, and again it was impossible to understand a word of what he was saying. Finally, Malinin could not take it any longer and put an end to this mutual torture: “**Don’t you** **try/endeavor**, Sirota, I can’t understand you anyway: your mouth got smashed .... There is only sound, no voice. You’ll be in hospital for a while – it will heal, but for now **don’t** **try**, don’t torture yourself” (...) [K. Simonov. Živye i mertvye]

out of our minds

TRY verbs (Divjak 2003, 2004; Divjak & Gries 2006)

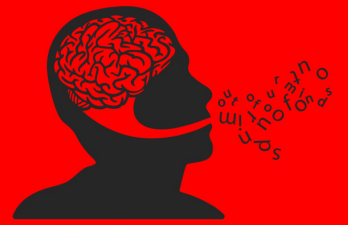
- 9 Russian verbs: **+inf** → *try*

  Probovat', *pytat'sja*, *starat'sja*, *silit'sja*, *norovit'*, *poryvat'sja*, *tscitsja*, *pyzit'sja*, *tyzit'sja*

- Studied in sample of 1585 usage ex

- Manually annotated
  - much of this can now be done automatically, depending on corpus

>>	то вре[релкий/Afpmsl]не была прошита в аппарате ... больше я не	пробовал	, если кто-то еще имел опыт - поделитесь ... FROL 2.02.2006
>>	- некоторые говорили, что можно нерестить 1x2 или 2x1, я	пробовал	, зря рыбсов мучал, как только разделил 1x1 дело пошло
>>	скрывал, шатался спокойно, пил себе пиво, какого в жизни не	пробовал	, и ни о чем не беспокоился. Как вскоре выяснилось,
>>	лучшие рукавицы для кайтсерфинга и виндсерфинга, которые я	пробовал	, катаясь в холодной воде Канады ». « Сначала я скептически
>>	всех? Ну вот и они пусть тебя поберегут. Уговоры ты уже	пробовал	, логику разъяснял - не доходит. Попробуй теперь по-другому,
>>	клуба &; quot; Барракуда &; quot;. Сам не	пробовал	, насколько там красивое дно не знаю. Ну, и, конечно же,
>>	нагрузки, принимать лучше в первой половине дня. Сам не	пробовал	, но говорят сбрасывают до 3 кг. в неделю без ущерба мышцам.
>>	, мечтал о путешествиях. В Африку, честно скажу, сбежать не	пробовал	, но любая самостоятельная вылазка в пригородный лес для нас
>>	проекты организовывать, с шоу -бизнесом там тоже что-то	пробовал	, но так, похоже, у него ничего и не получилось. Аксенцев в
>>	. 11. Бой - линейка. ИМХО - плохая идея с линейкой. Сам не	пробовал	, но уверен, что измерения таким способом ни к чему хорошему
>>	Chimay, Westmalle и Westvleteren. Последнего сорта я ещё не	пробовал	, потому, что не встречал нигде. Отличительной чертой
>>	, жарят и варят то, что ты никогда в жизни не видел и не	пробовал	, тут же читают свои газеты, смотрят и слушают по телевизору
>>	скажу, ни разу в жизни не ел, честно вам клянусь, даже не	пробовал	, что это такое, и вам не советую. (Аплодисменты.) Поэтому
>>	е vil: На тренажерах надрываешься, а толку - ноль: (Бегать	пробовал	- дыхалки не хватает. Спина, мать ее, как будто закостенела
>>	у нас тоже эта продукция в аптеках встречается. Я лично не	пробовал	- здоровье дороже. 8 долларов за 5 мл розы - это прикол: t
>>	эти &; quot; дела &; quot;, что я только не	пробовал	- и понижал дозу и меньше ангедрировал - результат один.
>>	... Читали классиков?) Harvi 2.02.2006 - 14:41 Zapper не	пробовал	- ничего сказать не могу, а аналогичную отечественную
>>	лучшее пиво - это темный « Старопрамен ». - К сожалению, не	пробовал	. - А я и не собираюсь больше пробовать, - заявила я, -
>>	. - Ну угости, а я тебе взамен дам то, чего ты 15 лет не	пробовал	. - Господи! Неужто пива привезла?! Утро. Похмелье. Мужик
>>	. Если вы хотите управлять один, будет еще сложнее. Я	пробовал	. - Как нужно построить структуру предприятия, чтобы оно
>>	внезапности? Доктор уныло пожал плечами: е знаю, мсье. Не	пробовал	. Ага, вот оно! Дверь приоткрылась, и в проеме появилась
>>	. Больше что-то не хочется. Ванька 20.10.20 11:52 Я тоже	пробовал	. В двух вариантах. Тоже не хочется ... Обмен энергией не
>>	по рукам. - Я твоего товара еще не видел, не взвешивал, не	пробовал	. Вдруг у тебя не героин, а подделка? Если у тебя то, что
>>	2005 года, 10:54 2 Aleko: Автоматом код подобрать даже не	пробовал	. Как прочитал описание процедуры - мне херовато стало. Тем
>>	я никогда, вообще никогда, не употреблял и даже не	пробовал	. Никогда, никакого. Никогда не курил. Я люблю спорт. Вот
>>	сигнала и фильтра сложения Cosmosa и МТИСа. Эфирные не	пробовал	. При этом берет все каналы и технологический COSMOSa. А
>>	это вслух. И уже будет не отнекаться. Я так раньше не	пробовал	. Я говорил только: &; quot; Господи, ты испытал меня
>>	. И не ДМТ. Хотя все вышеперечисленное Энди, несомненно,	пробовал	. Я придумал нечто. А потом отказался. Потому что я не
>>	кисло -соленым. Аж губы ест. Говорят, это хорошо к пиву, не	пробовал	. Я сумел съесть только половину пакетика, на пробу. Больше
>>	Буланка, пробовал когда-нибудь маслянки? — Н ... нет ... не	пробовал	. — Как? Ни разу не пробовал? — Ни разу ... — Вот так штука!
>>	? Ванька 20.10.20 12:43 В хорошем смысле. Он сначала	пробовал	... Потом ел ... Потом начинал жрать.. Потом ЖРАТЬ. И не
>>	этих самых носочков - внутрь, наружу, прямо. Кто может	пробовал	? А то еще хуже напортачу В джинсах - красота, в юбке - как
>>	! Или тебе мама прописала не расти в обратную сторону? А ты	пробовал	? А я? Вот весь летов об этом же - семь шагов за горизонт.
>>	. Задайте им вопросы, подобные следующим: &; quot; Что ты	пробовал	? Как это сработало? &; quot; &; quot; Что могу я
>>	поговорить с Евгением Владимировичем, с куратором? Кто -то	пробовал	? Костя нахмурился: – Был один … И нет его. Все встречи в п
>>	половину царства и принцессу в жёны &; quot;. Кто ни	пробовал	? ничего не получалось. И вот объявился Иван -дурак,
>>	маслянки? — Н ... нет ... не пробовал. — Как? Ни разу не	пробовал	? — Ни разу ... — Вот так штука! Хочешь, я тебя угощу? И, не
>>	, где-нибудь. Удачи 2. Coolmax - 1. Июнь 2004 8:02 Ты не	пробовал	MT-Textile 2? Если пробовал и у тебя получилось - чиркни на
>>	себя лучше. Кто -нибудь может это подтвердить? Я никогда не	пробовал	R5 поэтому не могу сравнивать, но то, что contra 16.5 даёт
>>	под руководством Чада, калифорнийца из-под Лос -Анджелеса,	пробовал	боди -сёрфить. Здорово. Нравится. Появилось желание взять
>>	вашем авто ...:) //26 MAXX 13 ianuarie, 10:29 //17 А ты не	пробовал	быть внимательным? Ты смотришь в сторону только когда на
>>	что там их делают очень хорошо, гораздо лучше чем те, что я	пробовал	в Израиле. Плюс к вышеописанному полному обслуживанию,
>>	Добавлено svetlyak, Вто, 18/10/2005 - 16:50 Юры,, а ты не	пробовал	взять у них программу того, что тебе не хватает, учить всё
>>	Yury, Вто, 18/10/2005 - 17:00 svetlyak пишет Юры,, а ты не	пробовал	взять у них программу того, что тебе не хватает, учить всё



Behavioral Profiles (Divjak 2004, 2010)

- Different from KWIC: narrow down/expand context window to “natural” unit of expression, i.e. sentence or clause

  > assumption: semantically most significant context is “natural” vicinity of word

- Multitude of properties

  > not known what does (not) convey meaning

ID tags (85) – within clause

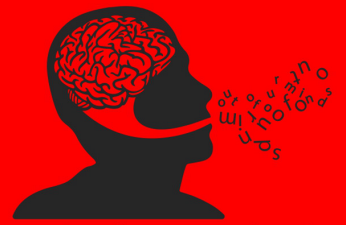
Subject: case + type of subject (9)

- animate (human being vs animal) vs inanimate (abstract vs concrete, man-made vs non-man made etc.)

ID tags (85) – within clause

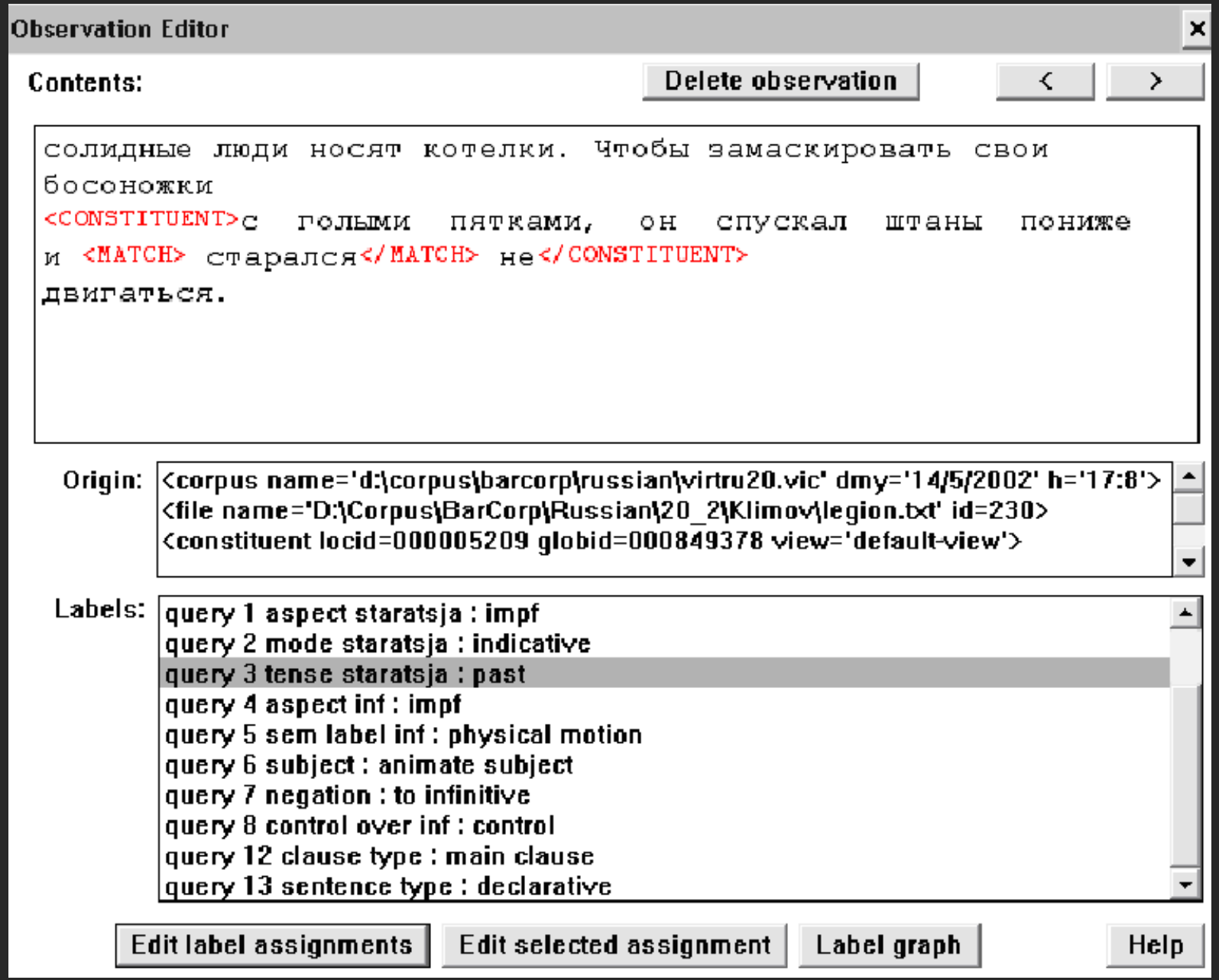
Infinitive: aspect + degree of control (low, medium, high) + type of action (15)

- physical action, perception, communication, intellectual activity, emotions etc.

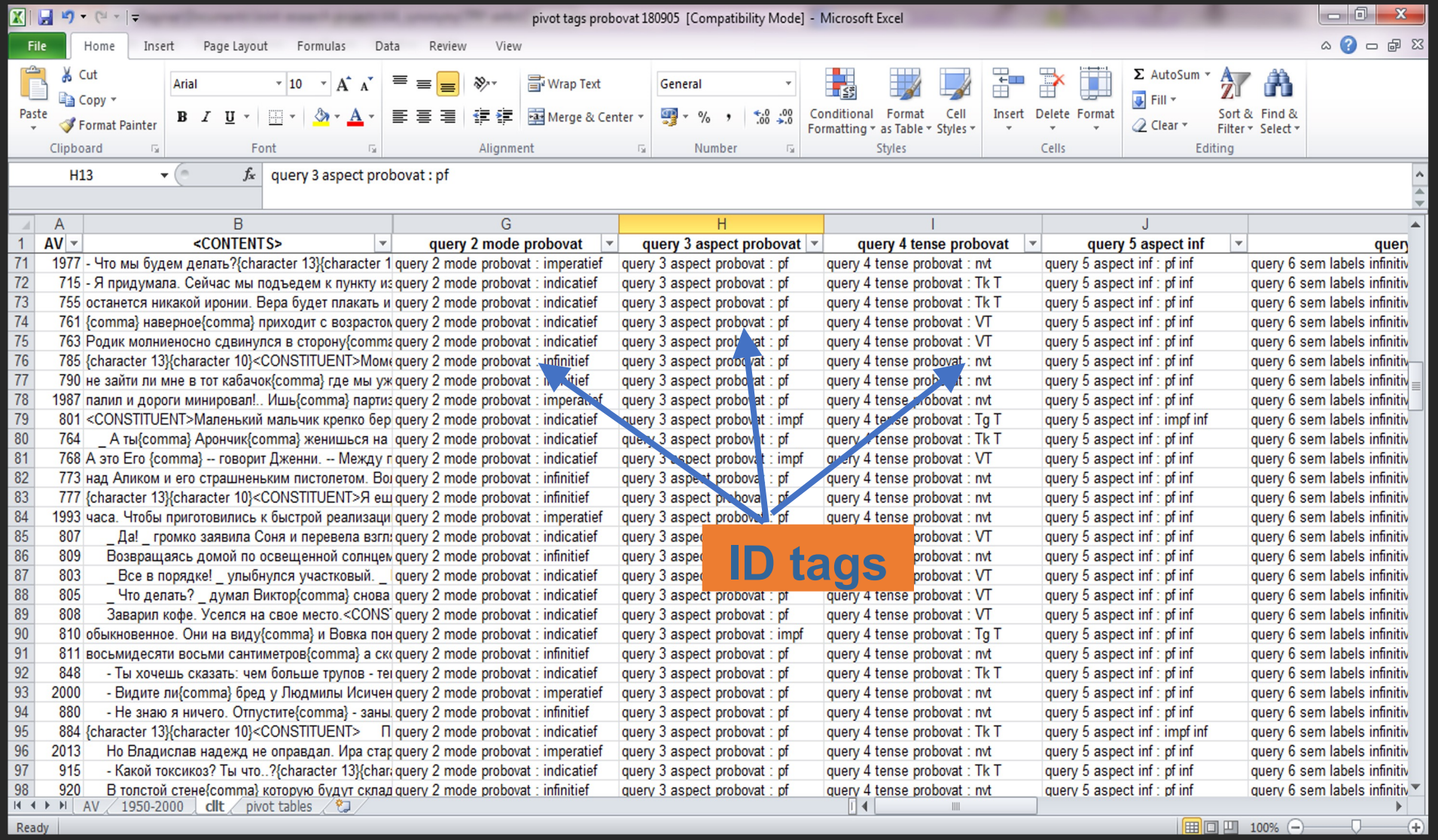


ID tags – within clause

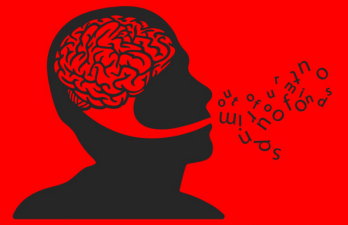
- **Finite verb**: aspect, mode, tense
  - e.g., *was trying* = past continuous
- **Optional elements**:
  - adverbs, particles and connectors, negation
  - e.g., had been trying *for a long time*
- **Clause/sentence type**:
  - In main vs subclause
    - e.g., *I tried* to explain BPs = in main clause
  - declarative vs imperative vs interrogative vs exclamative
    - e.g., *Try* to apply what you learn = imperative



To cover his sandals with open heels, he dropped his trousers a bit more and *tried* not to move.



	A	B	G	H	I	J	
1	AV	<CONTENTS>	query 2 mode probovat	query 3 aspect probovat	query 4 tense probovat	query 5 aspect inf	quer
71	1977	- Что мы будем делать?{character 13}{character 1	query 2 mode probovat : imperatief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
72	715	- Я придумала. Сейчас мы подъедем к пункту из	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : Tk T	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
73	755	останется никакой иронии. Вера будет плакать и	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : Tk T	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
74	761	{comma} наверное{comma} приходит с возрастом	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : VT	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
75	763	Родик молниеносно сдвинулся в сторону{comma	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : VT	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
76	785	{character 13}{character 10}<CONSTITUENT>Мом	query 2 mode probovat : infinitief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
77	790	не зайти ли мне в тот кабачок{comma} где мы уж	query 2 mode probovat :	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
78	1987	палил и дороги минировал!.. Ишь{comma} партиз	query 2 mode probovat : imperatief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
79	801	<CONSTITUENT>Маленький мальчик крепко бер	query 2 mode probovat : indicatief	query 3 aspect probovat : impf	query 4 tense probovat : Tg T	query 5 aspect inf : impf inf	query 6 sem labels infinitiv
80	764	_ А ты{comma} Арончик{comma} женишься на	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : Tk T	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
81	768	А это Его {comma} -- говорит Дженни. -- Между г	query 2 mode probovat : indicatief	query 3 aspect probovat : impf	query 4 tense probovat : VT	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
82	773	над Аликом и его страшненьким пистолетом. Во	query 2 mode probovat : infinitief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
83	777	{character 13}{character 10}<CONSTITUENT>Я еш	query 2 mode probovat : infinitief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
84	1993	часа. Чтобы приготовились к быстрой реализаци	query 2 mode probovat : imperatief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
85	807	_ Да! _ громко заявила Соня и перевела взгл	query 2 mode probovat : indicatief	query 3 aspec	probovat : VT	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
86	809	Возвращаясь домой по освещенной солнцем	query 2 mode probovat : infinitief	query 3 aspec	probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
87	803	_ Все в порядке! _ улыбнулся участковый. _	query 2 mode probovat : indicatief	query 3 aspec	probovat : VT	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
88	805	_ Что делать? _ думал Виктор{comma} снова	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : VT	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
89	808	Заварил кофе. Уселся на свое место.<CONS	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : VT	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
90	810	обыкновенное. Они на виду{comma} и Вовка пон	query 2 mode probovat : indicatief	query 3 aspect probovat : impf	query 4 tense probovat : Tg T	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
91	811	восьмидесяти восьми сантиметров{comma} а ск	query 2 mode probovat : infinitief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
92	848	- Ты хочешь сказать: чем больше трупов - те	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : Tk T	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
93	2000	- Видите ли{comma} бред у Людмилы Исичен	query 2 mode probovat : imperatief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
94	880	- Не знаю я ничего. Отпустите{comma} - заны	query 2 mode probovat : infinitief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
95	884	{character 13}{character 10}<CONSTITUENT> П	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : Tk T	query 5 aspect inf : impf inf	query 6 sem labels infinitiv
96	2013	Но Владислав надежд не оправдал. Ира стар	query 2 mode probovat : imperatief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
97	915	- Какой токсикоз? Ты что..?{character 13}{char	query 2 mode probovat : indicatief	query 3 aspect probovat : pf	query 4 tense probovat : Tk T	query 5 aspect inf : pf inf	query 6 sem labels infinitiv
98	920	В толстой стене{comma} которую будут склад	query 2 mode probovat : infinitief	query 3 aspect probovat : pf	query 4 tense probovat : nvt	query 5 aspect inf : pf inf	query 6 sem labels infinitiv



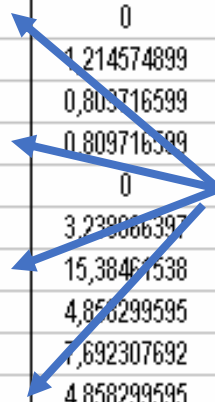
ID tag ~ Atkins (1987)

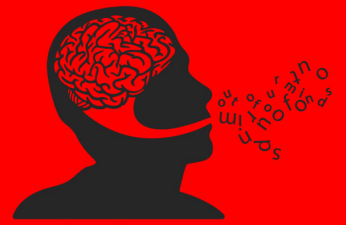
Label “ID tags” borrowed from Atkins (1987)

BUT only partial overlap:

> “syntactic or lexical markers in the citations which point to a particular dictionary sense of the word”
> (Atkins 1987: 24)

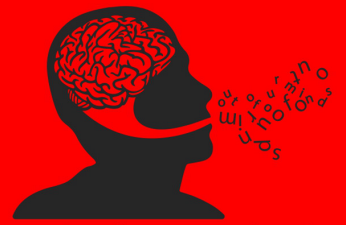
	A	C	D	E	F	G	H	I	J
1	VAR	pytat'sja	starat'sja	silit'sja	tscit'sja	tuzit'sja	pyzit'sja	norovit'	
8	7	0	0	0	0	0	2,040816327	0	0
9	8	1,619433198	1,612903226	1,652892562	0	5,660377358	1,020408163	0,840336134	6,4
10	9	0	0	0,41322314	0	0	1,020408163	0	0,8
11	10	0,4048583	0,403225806	1,652892562	4,166666667	9,433962264	7,142857143	1,680672269	4,4
12	11	0,4048583	0	0	4,166666667	1,886792453	2,040816327	0	0
13	12	0	0	1,239669421	0	1,886792453	0	0,840336134	6
14	13	0	0	1,239669421	0	0	1,020408163	0	3,2
15	14	1,214574899	0,403225806	0,826446281	5,555555556	3,773584906	4,081632653	0	0,4
16	15	0,809716599	0,403225806	0,41322314	4,166666667	5,660377358	6,12244898	0,840336134	0
17	16	0,809716599	0	0	0	1,886792453	1,020408163	0	0
18	17	0	0	0	0	0	1,020408163	0	0
19	20	3,238866397		6,198347107	2,777777778	7,547169811	3,06122449	2,521008403	8,4
20	21	15,38461538	9,274193548	6,611570248	12,5	22,64150943	11,2244898	8,403361345	20,4
21	22	4,858299595	5,64516129	0,826446281	2,777777778	5,660377358	3,06122449	7,56302521	9,2
22	23	7,692307692	7,258064516	6,198347107	4,166666667	7,547169811	5,102040816	42,85714286	26,4
23	24	4,858299595	1,612903226	3,719008264	2,777777778	0	1,020408163	4,201680672	4,4
24	25	11,74089069	11,29032258	2,892561983	9,722222222	24,5283		0	4,8
25	26	7,287449393	7,258064516	8,26446281	22,22222222	0		1,680672269	6
26	27	2,834008097	8,064516129	2,066115702	1,388888889	0	3,06122449	3,361344538	4,4
27	28	10,12145749	6,048387097	3,719008264	6,944444444	1,886792453	7,142857143	2,521008403	10
28	29	0,809716599	4,032258065	1,239669421	2,777777778	0	0	0,840336134	0,8
29	30	0	2,016129032	2,479338843	1,388888889	0	0	0	0
30	31	2,834008097	7,258064516	5,371900826	1,388888889	0	0	0,840336134	0,8
31	32	14,17004049	8,870967742	38,01652893	16,66666667	15,09433962	9,183673469	2,521008403	0
32	33	0,4048583	0,403225806	0	0	1,886792453	0	0	0
33	34	13,76518219	12,09677419	12,39669421	12,5	13,20754717	23,46938776	22,68907563	4,4
34	37	88,66396761	84,67741935	46,69421488	73,61111111	86,79245283	88,7755102	97,4789916	92,4
35	38	8,097165992	10,88709677	16,11570248	22,22222222	7,547169811	8,163265306	2,521008403	5,6
36	39	3,238866397	4,435483871	37,19008264	4,166666667	5,660377358	3,06122449	0	2
37	42	2,429149798	1,612903226	0	1,388888889	1,886792453	5,102040816	1,680672269	0





BP ~ Hanks (1996)

- “Brand name” borrowed from Hanks (1996) BUT
  - Hanks’ BP restricted to complementation patterns and semantic roles



Multivariate & multidimensional

- Is always multivariate:
  - each example annotated for *multitude* of parameters

  ← → work by Laura Janda and her team who've taken BPs apart again into grammatical profiles etc

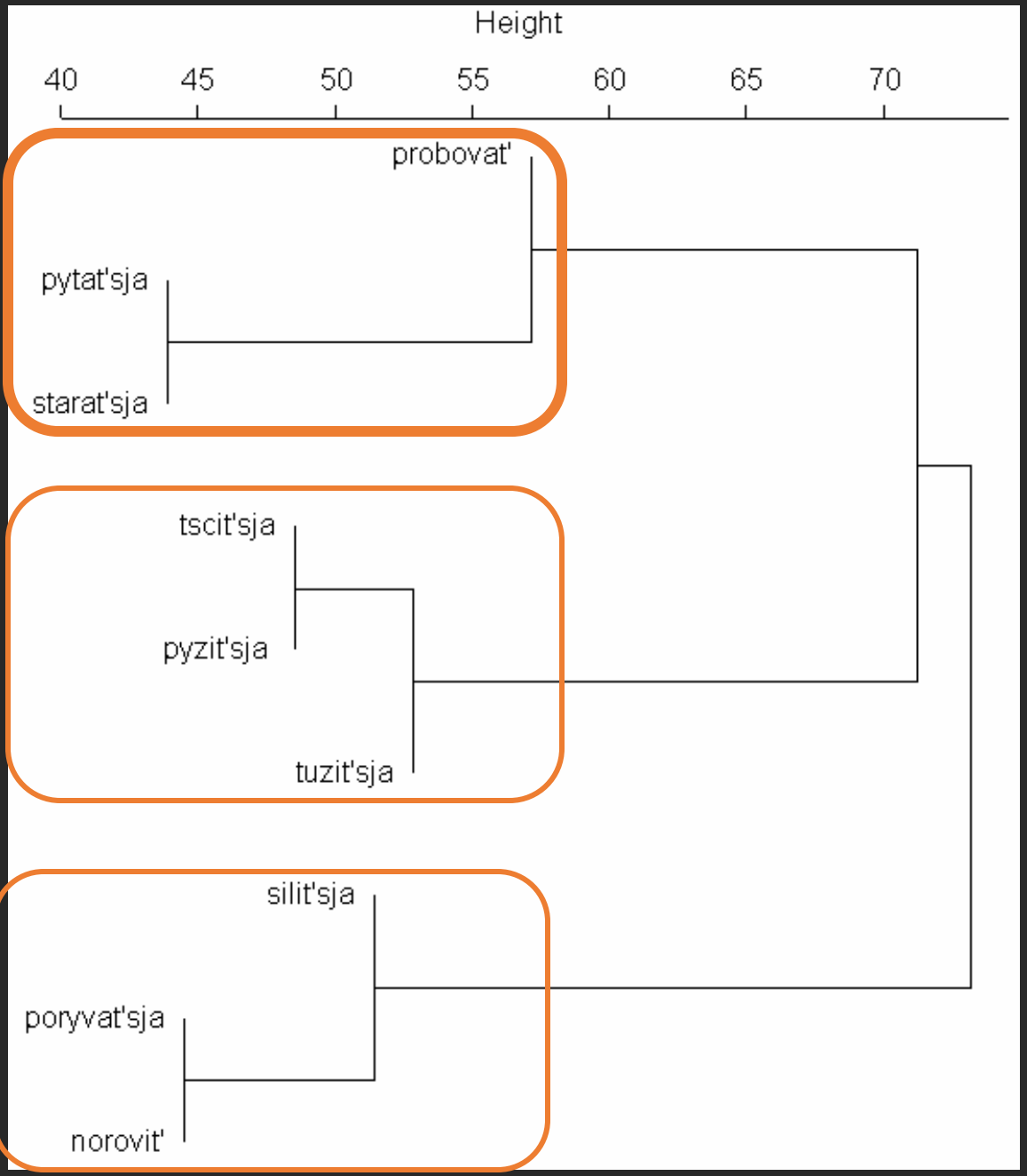
- Can also be multidimensional:
  - co-occurrence information for parameters preserved for each example

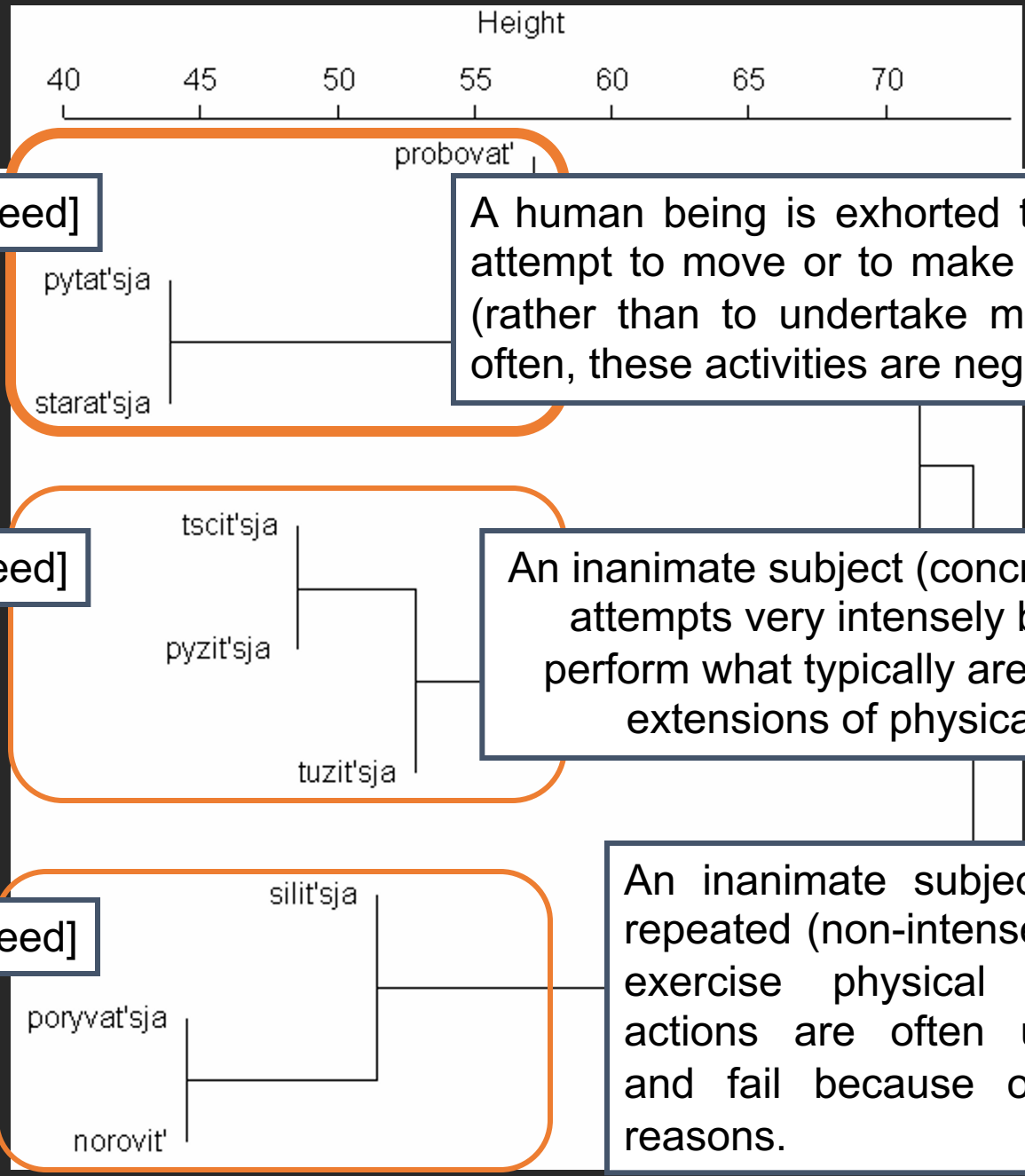
EXPLORE
(Divjak 2003; Divjak & Gries 2006)

- 9 verbs - 85 ID tags - 1585 examples
  - Co-occurrence table: +/-135 000 data points

- Find structure: cluster analysis [HAC]
  - Exploratory, hypothesis generating technique
    1. compares elements
    2. groups similar elements together

out of our minds





[you could succeed]

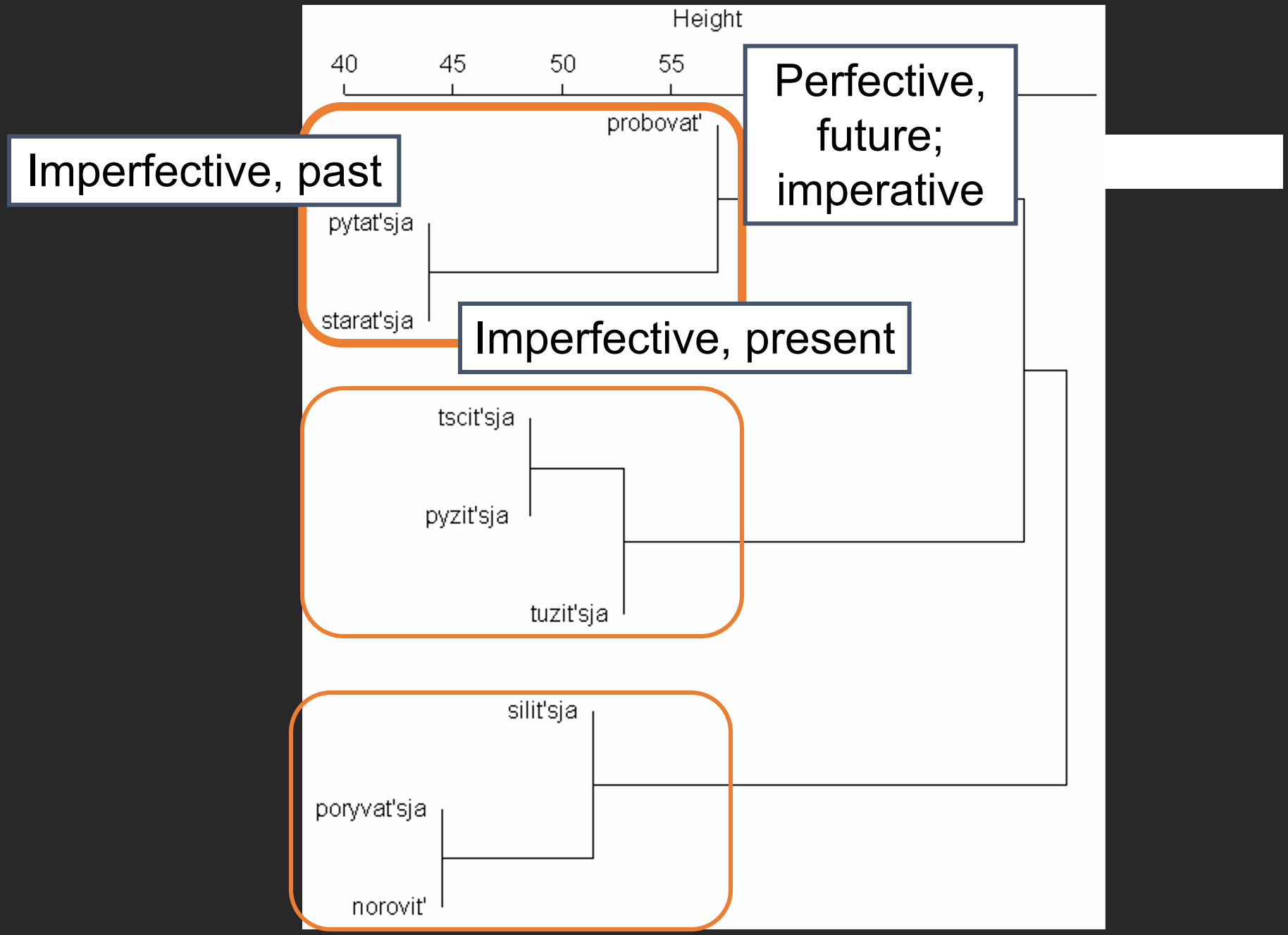
A human being is exhorted to undertake an attempt to move or to make someone move (rather than to undertake mental activities); often, these activities are negated.

[you can't succeed]

An inanimate subject (concrete or abstract) attempts very intensely but in vain to perform what typically are metaphorical extensions of physical actions.

[you won't succeed]

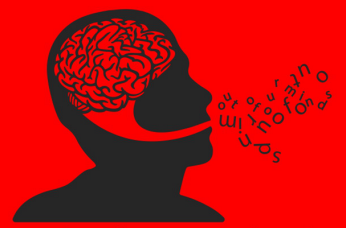
An inanimate subject undertakes repeated (non-intense) attempts to exercise physical motion; the actions are often uncontrollable and fail because of in-/external reasons.


Height
40
45
50
55
probovat'
pytat'sja
starat'sja
tscit'sja
pyzit'sja
tuzit'sja
silit'sja
poryvat'sja
norovit'
Imperfective, past
Perfective, future; imperative
Imperfective, present

Don’t You Try …

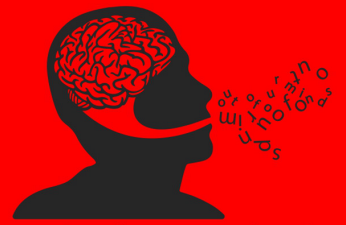
Но Сирота все еще **силился** **что-то сказать**, и снова невозможно было понять ни слова из того, что он говорил. Малинин наконец не выдержал и прекратил эту обоюдную муку: **Ты не** **старайся**, Сирота, все равно я не понимаю: у тебя рот разбитый … Звук и только, а голоса нет. В госпитале полежишь - восстановится, а сейчас **не** **пробуй**, не мучь себя (…) [К. Симонов. Живые и мертвые.]

But Sirota was still trying [intense attempt in vain] to say something, *and again it was impossible to understand a word of what he was saying*. Finally, Malinin could not take it any longer and put an end to this mutual torture: “Don’t you try [relatively intense, durative attempt that implies repetition], Sirota, *I can’t understand you anyway*: your mouth got smashed …. *There is only sound, no voice*. You’ll be in hospital for a while – it will heal, but for now don’t try [experimental, repeated attempt], don’t torture yourself” (…)



Behavioral Profiles (Divjak 2004, 2010)

- Labels
  - **naïve**: require no “linguistic” insight/analysis
  - exhaustive
    - ! bottom-up established, not pre-set
  - ! could do without
    - labels used to overcome data sparseness

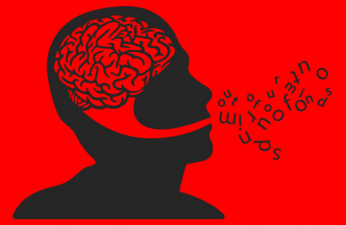


Core

- Morphology - Grammatical preferences
  - TAM, voice, number, person …
- Syntax - Constructional preferences
  - Properties of each construction: main/sub, declarative/interrogative/imperative
- Semantics - Semantic preferences
  - obligatory *and* optional slots

! Flexible

  - BP can easily be expanded to contain information about text type, writer gender, regiolect, font size …



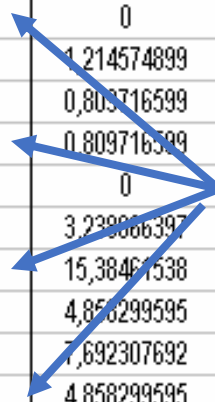
Statistical analysis

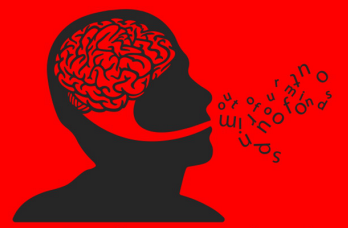
- Provides data that can be modelled in a variety of ways (<> Gries 2012)
  - Depends on research question (Divjak 2010)
  - Need to adapt dataset format according to requirements of technique
    - Cluster Analysis ~ vectors (BehavioralProfiles 1.0)
    - Regression: retain link to other properties in same sentence ~ conditional probabilities

NLP implementations

- Range of models
  - Vector space models: LSA
  - Probabilistic Topic Models
- Issues typically discussed:
  - Context: size of unit varies from 2-word window to text
    - Larger ~ topical information → information retrieval
    - Smaller ~ lexical semantic competence

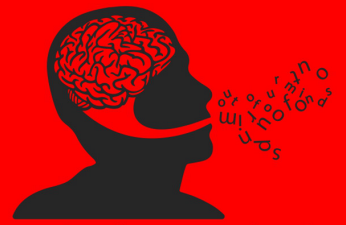
	A	C	D	E	F	G	H	I	J
1	VAR	pytat'sja	starat'sja	silit'sja	tscit'sja	tuzit'sja	pyzit'sja	norovit'	
8	7	0	0	0	0	0	2,040816327	0	0
9	8	1,619433198	1,612903226	1,652892562	0	5,660377358	1,020408163	0,840336134	6,4
10	9	0	0	0,41322314	0	0	1,020408163	0	0,8
11	10	0,4048583	0,403225806	1,652892562	4,166666667	9,433962264	7,142857143	1,680672269	4,4
12	11	0,4048583	0	0	4,166666667	1,886792453	2,040816327	0	0
13	12	0	0	1,239669421	0	1,886792453	0	0,840336134	6
14	13	0	0	1,239669421	0	0	1,020408163	0	3,2
15	14	1,214574899	0,403225806	0,826446281	5,555555556	3,773584906	4,081632653	0	0,4
16	15	0,809716599	0,403225806	0,41322314	4,166666667	5,660377358	6,12244898	0,840336134	0
17	16	0,809716599	0	0	0	1,886792453	1,020408163	0	0
18	17	0	0	0	0	0	1,020408163	0	0
19	20	3,238866397		6,198347107	2,777777778	7,547169811	3,06122449	2,521008403	8,4
20	21	15,38461538	9,274193548	6,611570248	12,5	22,64150943	11,2244898	8,403361345	20,4
21	22	4,858299595	5,64516129	0,826446281	2,777777778	5,660377358	3,06122449	7,56302521	9,2
22	23	7,692307692	7,258064516	6,198347107	4,166666667	7,547169811	5,102040816	42,85714286	26,4
23	24	4,858299595	1,612903226	3,719008264	2,777777778	0	1,020408163	4,201680672	4,4
24	25	11,74089069	11,29032258	2,892561983	9,722222222	24,5283		0	4,8
25	26	7,287449393	7,258064516	8,26446281	22,22222222	0		1,680672269	6
26	27	2,834008097	8,064516129	2,066115702	1,388888889	0	3,06122449	3,361344538	4,4
27	28	10,12145749	6,048387097	3,719008264	6,944444444	1,886792453	7,142857143	2,521008403	10
28	29	0,809716599	4,032258065	1,239669421	2,777777778	0	0	0,840336134	0,8
29	30	0	2,016129032	2,479338843	1,388888889	0	0	0	0
30	31	2,834008097	7,258064516	5,371900826	1,388888889	0	0	0,840336134	0,8
31	32	14,17004049	8,870967742	38,01652893	16,66666667	15,09433962	9,183673469	2,521008403	0
32	33	0,4048583	0,403225806	0	0	1,886792453	0	0	0
33	34	13,76518219	12,09677419	12,39669421	12,5	13,20754717	23,46938776	22,68907563	4,4
34	37	88,66396761	84,67741935	46,69421488	73,61111111	86,79245283	88,7755102	97,4789916	92,4
35	38	8,097165992	10,88709677	16,11570248	22,22222222	7,547169811	8,163265306	2,521008403	5,6
36	39	3,238866397	4,435483871	37,19008264	4,166666667	5,660377358	3,06122449	0	2
37	42	2,429149798	1,612903226	0	1,388888889	1,886792453	5,102040816	1,680672269	0





Embeddings

- Embeddings ~ a continuous variant of categorical BPs
- BPs are
  - Discrete: encoding presence versus absence
  - sparse > manual annotation limits the number of features that can be included
- Embeddings are
  - real-valued and dense,
  - typically one or two orders of magnitude larger than standard BPs
  - package world knowledge (Grand et al. 2022), not encoded by BPs



Embeddings

! BPs have been shown to surpass LLMs in detecting shifts in meaning (Guilianelli et al. 2022)

→superior ability to capture fine-grained morphological and syntactic signals

→ excellent point of contact between usage-based, corpus-based linguistics and LLMs

→ excellent way to probe what LLMs are capturing and what linguistic theory is missing

out of our minds

Thank you!

LEVERHULME TRUST

UNIVERSITY OF BIRMINGHAM