

Семантические карты

Исследования лингвоспецифической лексики

[Sharoff 2002]

“Meaning as use: exploitation of aligned corpora for the contrastive study of lexical semantics”

Dictionaries (mono- and multilingual) are a good repository of information on meanings that can be expressed by a word and on its possible translations. However, dictionaries cannot list all possible contexts, in which a word can be used and which can influence its translation into another language. Moreover, dictionaries are designed for helping a human reader (with some expertise) to understand or translate a word in a specific context, and not for providing linguistic analysis of lexical semantics or helping in machine translation or generation.

[Sharoff 2002]

Unlike logic-based theories of meanings, which assume that a word denotes a concept (a word has a meaning as an entry in a dictionary), functional theories of meaning assume that the meaning of a word is the function of its purposeful use in communication.

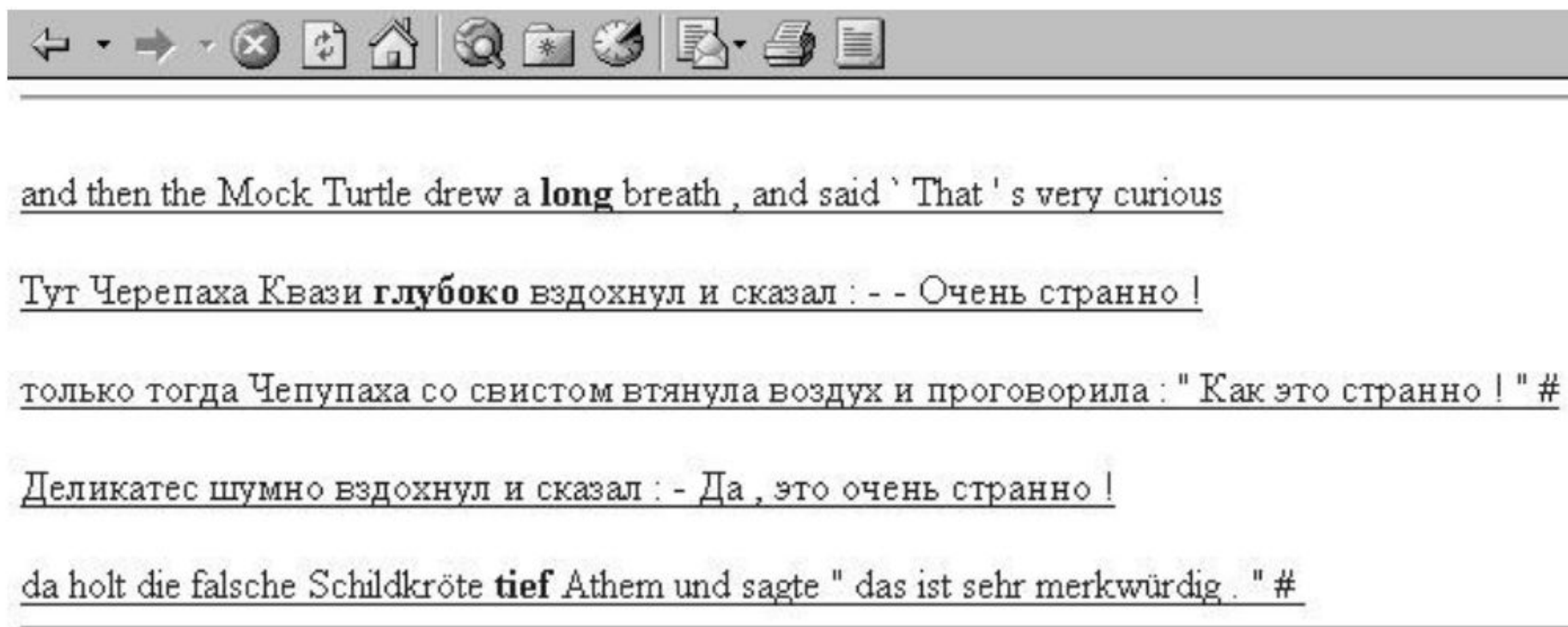


Figure 1. An output from processing the query: /<w id="(\\S*)" lemma="long" pos="adj"/

```

<entry key="deep" lang="en">
  <sense n="1" resp="cobuild"><def>If something is deep, it extends a long way down from the ground
    or from the top surface of something.</def>
    <semref type="ant" target="shallow"/>
    <eg resp="alice"><q id="alice.11">she found herself falling down a very deep well. </q>
      <trans><tr id="alice-d.17" >она начала падать, словно в <semref type="trans">глубокий</semref>
        колодец.</tr>
        <tr id="alice-g.13">sie fiel, wie es schien , in einen <semref type="trans">tiefen</semref>
          Brunnen.</tr></trans></eg>
    <eg resp="cobuild"><q>Den had dug a deep hole in the centre of the garden.</q></eg>
    <fts>spatiotemporal big neutral-interpersonal vertical non-emphasized spatial-lex depth-size</fts>
  </sense>

  <sense n="11" resp="cobuild"><def>A deep sound is low in pitch.</def>
    <semref type="ant" target="high"/>
    <eg><q id="alice.1387">said the Mock Turtle in a deep, hollow tone. </q>
      <trans><q id="alice-d.1913">проговорил Квази <semref type="trans">глухим</semref> голосом</q>
        <q id="alice-n.1502">ответила Чепуха <semref type="trans">глубоким</semref> голосом</q>
        <q id="alice-g.1224">sprach die falsche Schildkröte mit <semref type="trans">tiefer
          </semref> Stimme</q></trans></eg>
  </sense>

```

Figure 2. The structure of a lexical entry for deep

[Добровольский 2014]

В русском языке для выражения положительного vs. отрицательного согласия-несогласия есть только два слова: *да* и *нет*. Причем в нормированной речи *да* употребляется только в случае «положительного согласия» (i), а во всех остальных случаях – *нет*. Ср.:

(i) – *Будешь?* – ***Да***, буду;

(ii) – *Будешь?* – ***Нет***, не буду;

(iii) – *Не будешь?* – ***Нет***, буду;

(iv) – *Не будешь?* – ***Нет***, не буду

Знаменитый предвыборный лозунг Обамы *Yes, we can!*, обсуждаемый в [Dobrovol'skij, Levontina 2012] в связи с особенностями употребления русского *нет* в контекстах такого типа.

*They say, we are not ready, we cannot. **Yes**, we can!*

Нам говорили, что мы не готовы, что мы не можем.

****Да**, мы можем! при норме **Нет**, мы можем!*

Попробуем зарисовать

(i) – Будеишь? – Да , буду; Yes Ja	(ii) – Будеишь? – Нет , не буду; No Nein
(iii) – Не будеишь? – Нет , буду; Yes/No Nein	(iv) – Не будеишь? – Нет , не буду Yes/No Doch

[Добровольский 2014]

Соответствующие параллельные тексты представляют собой обозримый фрагмент корпуса, однородный как с точки зрения авторского стиля, так и выбранных переводчиком стратегий (все рассказы переведены на английский Дмитрием Набоковым):

1. реакцию на невысказанное: реакцию на невысказанное (– *Что мы тогда пели? **Нем**, это вспомнить невозможно* [Timberlake 2004: 465].):

“What we’ll really do,” she said, “is get you ready for your friend Dr. Valentini.” (E. Hemingway. A Farewell to Arms

– *Нем*, сейчас мы займемся другим делом, – сказала она, – нужно приготовить тебя к приходу твоего друга доктора Валентини.

2. усиление высказывания, подчеркивание важности сообщения

I tell you this war is a bad thing. (E. Hemingway. A Farewell to Arms)

Нем, в самом деле, скверная штука война.

[Добровольский 2014]

3. удвоения отрицания

Selbst dort zu arbeiten und zu unterhandeln, hat sich wenigstens für mich als ganz unmöglich erwiesen.
(F. Kafka. Der Prozess)

Действовать там самому, самому вести переговоры – *нет*, мне это оказалось совершенно не под силу.

4. ввода самокомментария

Мало сказать: он служил ревностно, *нет*, он служил с любовью. (Н.В. Гоголь. Шинель)
Akaki Akakiewitsch diente mit Eifer, *doch* das ist noch nicht das Wort: er diente mit Liebe.

5. Обращение к параллельным корпусам позволило выявить еще один тип контекстов, в которых русскому *нет* в немецком языке соответствует не *nein*, а какое-либо другое выражение. Это так называемое приимперативное *нет*

Ich muß schon gehen.“ „Halt, warte!“ (Th. Mann. Der Zauberberg)
– Мне пора. – *Нет*, подожди, постой!

Дискурсивное слово “вот”: RU - EN

Попробуем классифицировать употребления из корпуса, если останется время

→ Больше о дискурсивных единицах

Pragmatic Markers, Discourse Markers and Modal Particles

New perspectives

Edited by
Chiara Fedriani
Andrea Sansò

Как устроена работа?

1. В первой секции обсуждаются дискурсивные и прагматические маркеры
2. Во второй – модальные частицы
3. Далее речь идет о диахроническом аспекте развития указанных выше элементов
4. Социолингвистический аспект

[Just a suggestion: just/e in French and English \[Kate Beeching\]](#)

This article aims to illustrate what a contrastive perspective, combined with sociolinguistic and corpus approaches, can bring to the investigation of the evolution of pragmatic functions which emerge through language contact (Ingham 2012a) and pragmatic borrowing (Andersen 2014).

Adverbial usages of *just/e* include:

a. Temporal and spatial precision

<i>Just</i>	<i>juste</i>
'just before', 'just after', 'just following', 'just coming up to five', 'just beside'	'juste auparavant', ('just before') 'juste à côté', ('just beside') 'juste en face' ('just opposite'), 'juste en dessous' ('just below')

b. Recent past

<i>Just</i>	<i>juste</i>
'we've just filed', 'I've just described', 'Ros was just saying'	'On vient (juste) de...' ('we've just...')

c. Restrictive

<i>Just</i>	<i>juste</i>
(only) 'it'll just be for one product', 'just a list of images', 'just a third'; (simply) 'you can just apply every now and again'	'j'étais partie sans tente juste un sac à dos', ('I had left without a tent, just a sleeping bag') 'des lecteurs qui lisent juste le texte qui regardent très peu les images', ('readers who read just the text who don't look at the pictures very much'), 'il aura juste à faire la demande' ('he'll just have to ask')

Pragmatic usages of *just/e* include:

a. Minimisation of additions, questions, suggestions, criticisms and requests

<i>Just</i>	<i>juste</i>
'just a word', 'just a bit more', 'just a suggestion', 'can I just ask', 'could you just sort of mark that in your diary.'	'tu as juste à brancher ta machine', ('you just have to plug in the machine') 'juste une remarque' ('just a comment'), 'je veux juste répondre' ('I just want to reply'), 'je voulais juste vous demander donc quelques précisions' ('I just wanted to ask you a few details'), 'j'peux juste vous faire signer' ('can I just get you to sign')

b. Minimisation of assertions

<i>Just</i>	<i>Juste</i>
'I just think that'; 'I was just wondering...'	

c. Minimisation of reasoning or justification

<i>Just</i>	<i>juste</i>
'it's just that we had this issue'	'c'est juste que ma sœur ... est pas mariée là' ('it's just that my sister..isn't married')

d. Intensifying

<i>Just</i>	<i>juste</i>
'I just don't know' ; 'just amazing'	

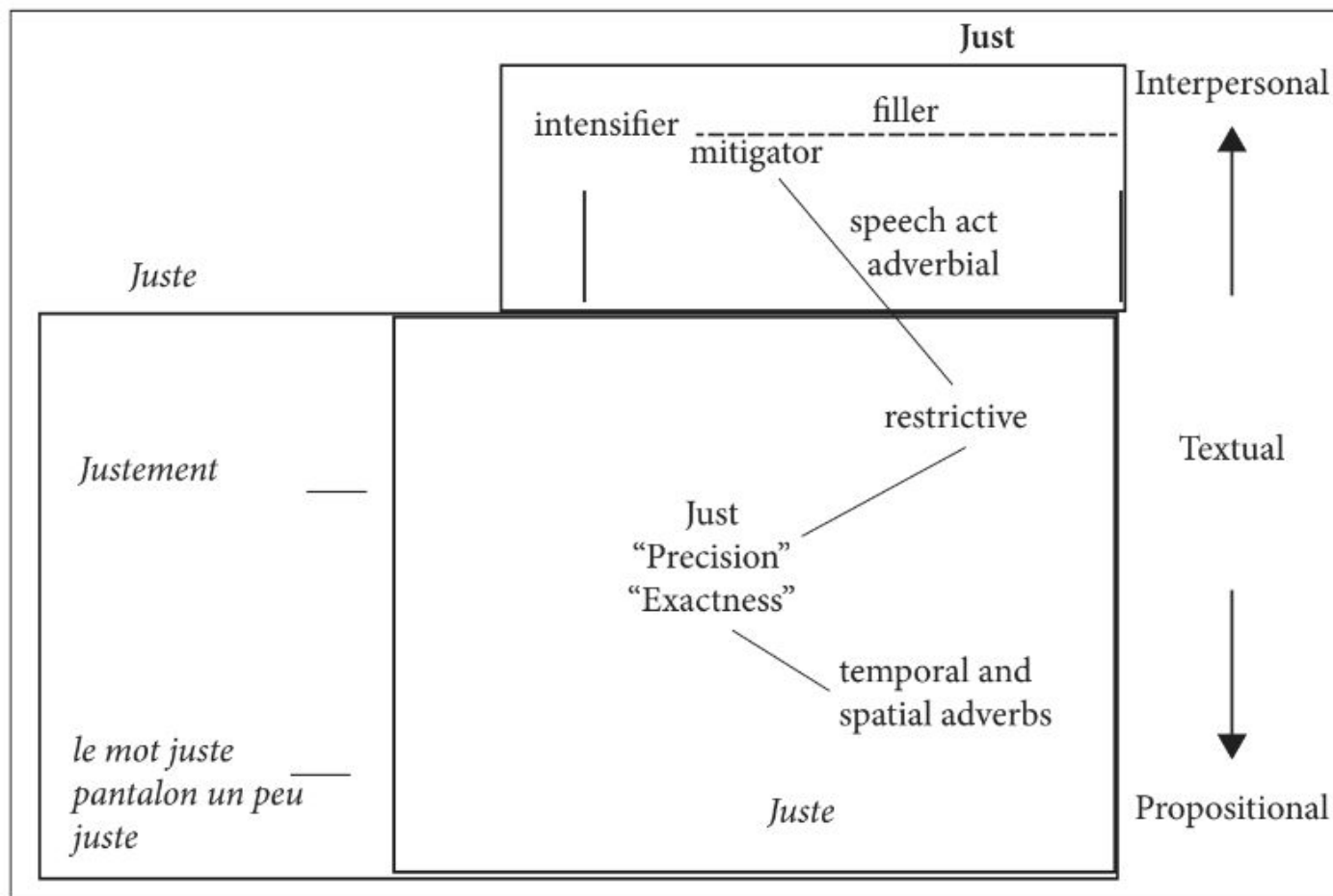


Figure 5. Semantic map of *just/e(ment)*

Вопросы филогенетического родства

Преимущества пословного выравнивания также открывают возможности для сравнения языков с целью определения их филогенетического родства.

Такое исследование представлено в работе [\[Mayer, Cysouw 2012\]](#)

Данные - тексты религиозных брошюр на 146 языках

При определении родства авторы основываются на частотах совпадения слов в цепочках выравнивания. Слова из разных языков часто встречающиеся в рамках одного выравнивания образуют кластеры

[Mayer, Cysouw 2012]

Определение сходства происходит следующим образом: слова из разных языков, попавшие в один кластер, т.е. имеющие высокую частоту соответствия друг другу в выравниваниях, считаются родственно близкими. Этот подход авторы предлагают использовать как альтернативу спискам Сводеша.

1. ису_{bl} jesus_{en} fia_{ew} yesu_{ew} gesù_{mt} jesús_{es}
jesus_{de}
2. кой_{bl} who_{en} min_{mt} wer_{de}
3. regieren_{de}
4. управлява_{bl} adu_{ew} jañkem_{mt} gobernarán_{es}
5. amekawoe_{ew} quiénes_{es}
6. ще_{bl} will_{en} se_{mt} wird_{de}
7. c_{bl} with_{en} con_{es} mit_{de}
8. kple_{ew}
9. ma_{mt}
10. rule_{en}

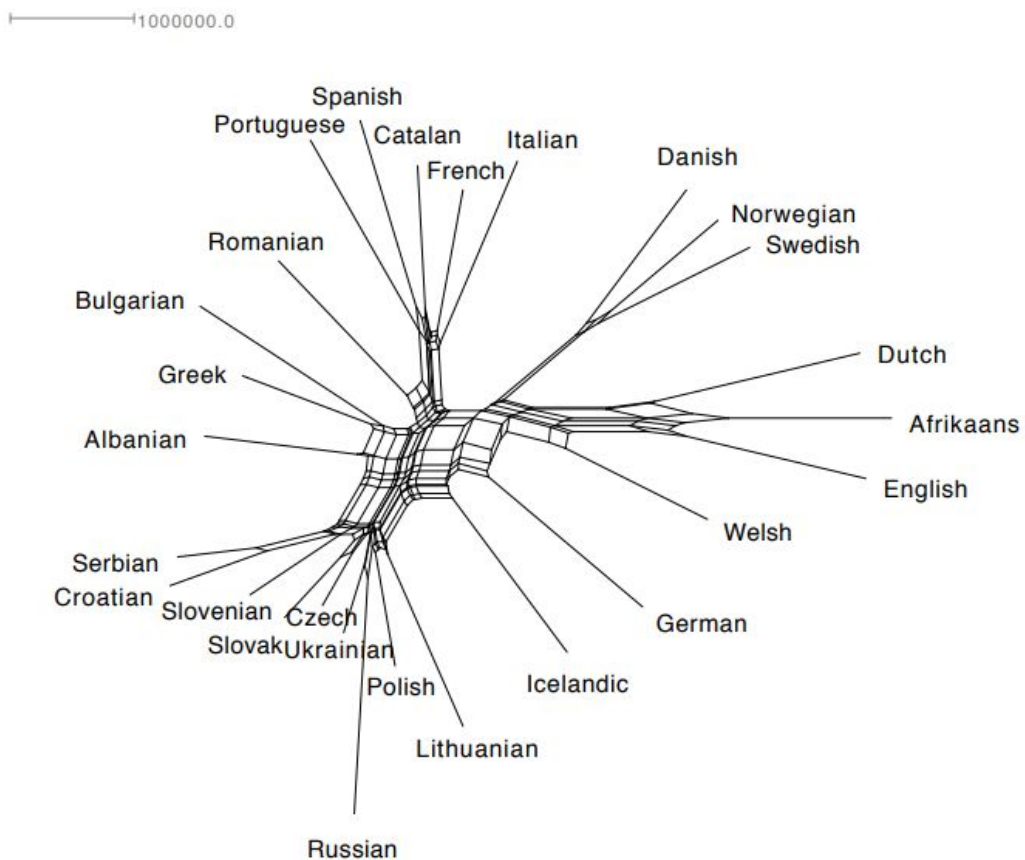


Figure 2: NeighborNet (created with SplitsTree, Huson and Bryant (2006)) of all Indo-European languages in the sample

Предпосылки создания датасета для лексико-грамматического исследования

Создание датасета

Проведение количественного лексико-грамматического исследования требует не только наличия хорошо сконструированного корпуса, но еще и датасета, построенного на базе этого корпуса

Опираясь на гипотезу об устройстве языкового явления, исследователь извлекает из корпуса релевантную информацию и структурирует ее.

Рассмотрим подробнее, как это реализуется на практике.

Формат датасета

Датасет представляет собой данные в особом формате, включающие в себя всю релевантную и необходимую для исследования информацию. Датасет содержит: а) примеры предложений с употреблениями исследуемого явления; б) тэги разновидностей способов выражения исследуемого явления — это может касаться разного оформления, например, каким предлогом вводится дополнение или какая форма глагола используется; в) значения различных признаков, например, семантический класс глагола. К данным, упорядоченным таким образом, можно применять различные статистические методы анализа и сложной визуализации

→ Имея датасет, можно переходить к исследованию

Методы, позволяющие анализировать
мультязычный материал

Методы квантитативного исследования

1. Факторный анализ
2. Сетевой анализ
3. Построение семантических карт

Факторный анализ

В своей работе [Gries 2001] С. Грис исследует влияние различных факторов на выбор места частицы фразового глагола (John picked up the book. vs. John picked the book up). В эксперименте задействованы как чисто формальные признаки, например, длина прямого дополнения, так и грамматические — одушевленность прямого дополнения.

Объектом классификации является способ реализации фразового глагола с его прямым дополнением.

Стефан Грис, “Quantitative corpus linguistics with R: A practical introduction”.

Факторный анализ

[Bresnan, Ford 2010], посвященная дативной альтернации, то есть различному кодированию дативного дополнения:

- (1) a. Who gave that wonderful watch to you? (prepositional (to-)dative)
- b. Who gave you that wonderful watch? (double-object construction)

В качестве признаков используются характеристики участников ситуации, например, является ли тема/реципиент местоимением, число, одушевленность и определенность темы/реципиента. Факторный анализ позволяет установить взаимосвязь между обозначенными в эксперименте признаками и способом выражения

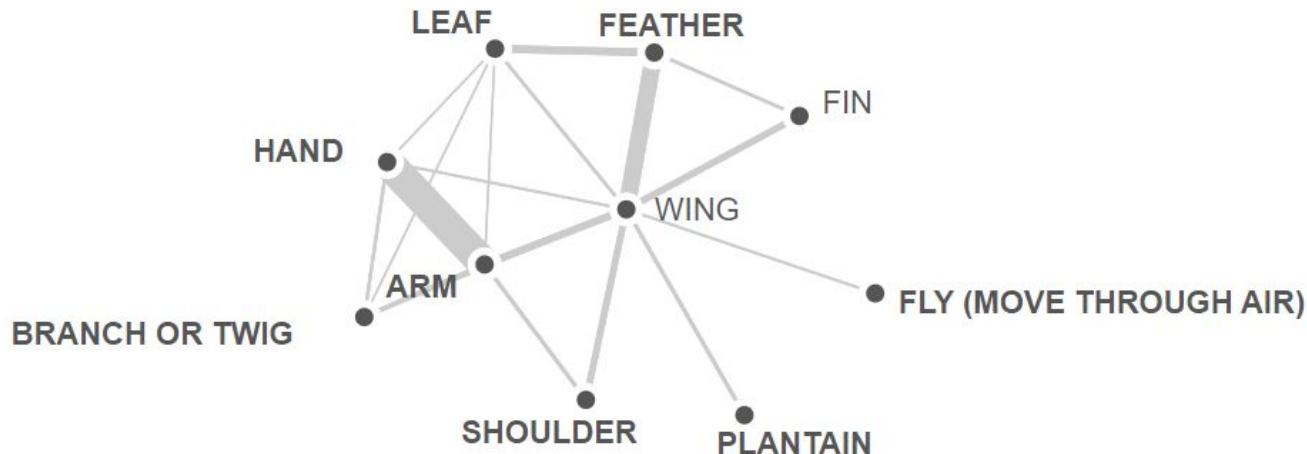
Факторный анализ

На основе признаков, используемых в исследовании, строится предсказательная модель, характеристики которой указывают на вклад признаков в предсказание выбора конструкции. В результате обучения моделей регрессии каждому признаку приписывается положительный или отрицательный коэффициент, а также значение p-value, которое указывает на значимость признака: если значение меньше 0,05, то признак является значимым

<https://agerasimova.com/ru/papers/>

Сетевой анализ

Модель графа состоит из узлов и ребер, указывающих на наличие связи между узлами, и узлы, и ребра имеют количественные характеристики, которые могут указывать, например, на количество связей у узла или частоту совместной встречаемости элементов, соответствующих узлам, друг с другом.



Concept WIDE

Colexified concepts:

Links	Concept
-------	---------

64	BIG
----	-----

17	THICK
----	-------

11	FLAT
----	------

5	SHORT
---	-------

4	LONG
---	------

Links	Concept
-------	---------

63	TAKE
----	------

28	GET
----	-----

25	SELL
----	------

14	CARRY
----	-------

13	PAY
----	-----

10	SEIZE
----	-------

9	TRADE OR BARTER
---	-----------------

8	HOLD
---	------

6	PICK UP
---	---------

5	WANT
---	------

4	CRY
---	-----

4	CATCH
---	-------

4	ASK (REQUEST)
---	---------------

3	CHANGE
---	--------

3	SHOUT
---	-------

Concept BUY

Links	Concept
63	TAKE
28	GET
25	SELL
14	CARRY
13	PAY
10	SEIZE
9	TRADE OR BARTER
8	HOLD
6	PICK UP
5	WANT
4	CRY
4	CATCH
4	ASK (REQUEST)
3	CHANGE
3	SHOUT

Сетевой анализ

Помимо того, что модель графа является удобным инструментом визуализации, детальное изучение его устройства позволяет получить больше информации об исследуемом явлении и систематизировать ее. Этот метод анализа дает возможность посмотреть на данные иначе, чем на семантической карте, и получить более подробное представление о том, какие языковые средства выражения регулярно используются в качестве переводных эквивалентов друг друга. Таким образом, результаты, которые мы получаем при использовании сетевого анализа, дополняют те, что мы получаем, применяя другие методы.

Бонч-Осмоловская А. А. Нестеренко Л. В. Сети как инструмент поиска и находок в мультязычных параллельных корпусах. под ред. Д. А. Рыжова [и др.]., Лабиринт, 2018.С. 305–320.

Построение семантических карт

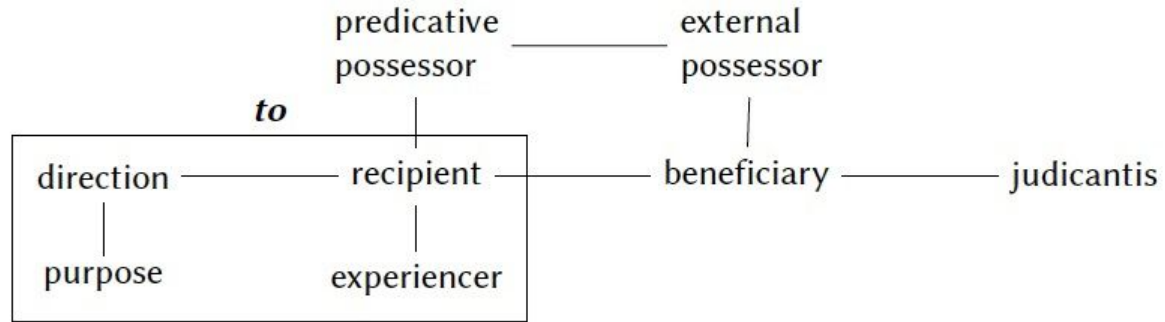
Семантическая карта (semantic map)

графически изображает значения/функции из определенного концептуального пространства (conceptual space), предсказывая возможную дистрибуцию конкретно-языковых средств кодирования данных функций.

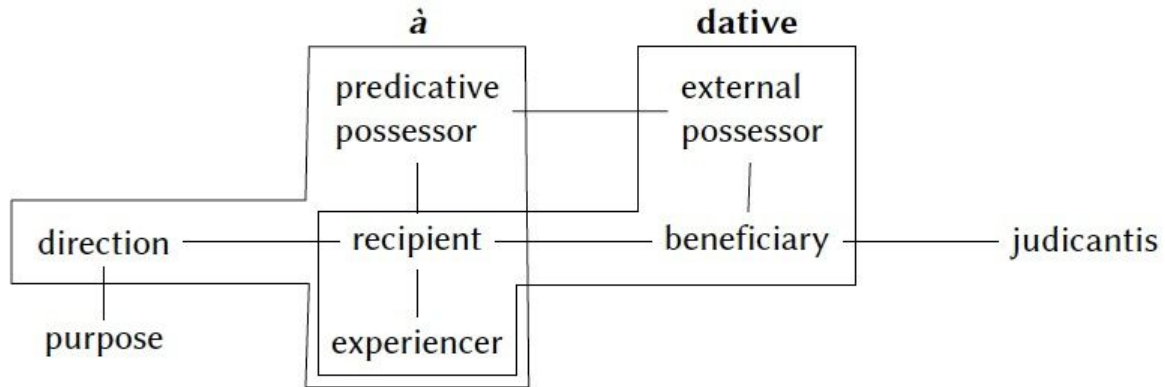
- хотя карта называется «семантической», на ней изображены не столько собственно значения, сколько «функции» (или: «употребления»)
- как правило, речь идет о функциях, кодируемых грамматическими формами

Идея использования семантической карты как модели для межъязыкового анализа была разработана и предложена Ллойдом Андерсеном [Anderson 1974, 1986, 1987].

- Английский



- Французский



Классический пример семантической карты:

«дательные» функции

- английский

recipient Eve gave the apple **to** Adam.

direction Goethe went **to** Leipzig as a student.

experiencer This seems outrageous **to** me.

purpose I left the party early **to** go home.

- французский (à)

recipient Donne le livre **à** moi.

direction Je vais **à** Moscou.

experiencer Cela plaît **à** mon frère.

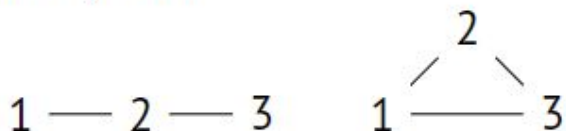
purpose — (pour + inf.)

Принципы построения

- Узлы — ярлыки для функций
- Области — дистрибуция конкретно-языковых показателей
- Функция вносится на карту, если хотя бы пара языков различается по средству выражения данной функции
 - ▶ с одной стороны, вовсе не обязательно между значениями должна быть какая-то прямая связь
 - ▶ с другой стороны, в конкретном языке их, возможно, не следует разделять:

хабитуалис — прогрессив — будущее

- Область одного показателя на карте **не должна быть разрывной**
- Чем меньше на карте линий, тем она более ограничительна и интересна:



Принцип смежности (Татевосов 2002: 33)

Любые два значения на семантической карте кодируются одним и тем же грамматическим средством тогда и только тогда, когда это же средство кодирует все значения, расположенные между ними.

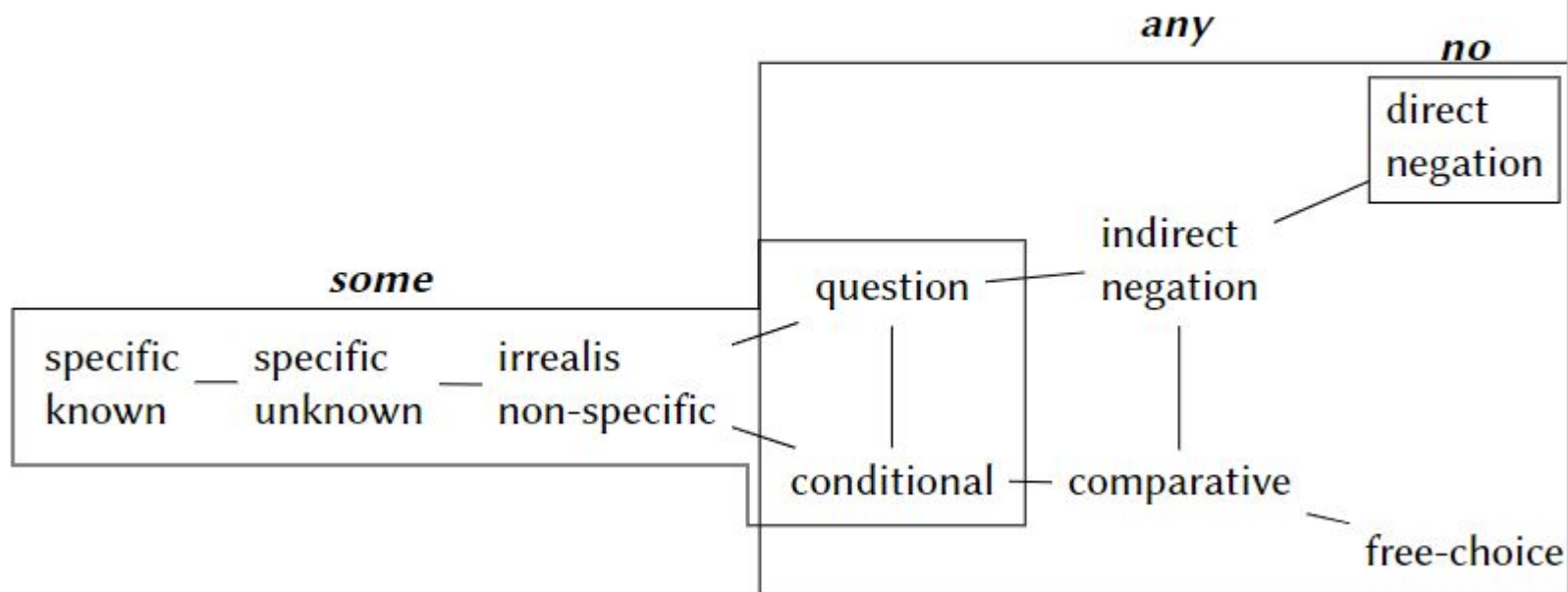
Карта для лексического поля 'дерево' / 'лес' (Хаспельмат, на основе одного из примеров у Ельмслева)

- в результате сравнения полисемии в нескольких европейских языках пять значений можно упорядочить определенным образом:
 - ▶ дерево – древесина – дрова – роща – лес
- если слово выражает значения А и С, то выражает и промежуточное между ними В

	дерево	древесина	дрова	роща	лес
немецкий	<i>Baum</i>	<i>Holz</i>		<i>Wald</i>	
датский		<i>trae</i>		<i>skov</i>	
французский	<i>arbre</i>		<i>bois</i>		<i>forêt</i>
испанский	<i>árbol</i>	<i>madera</i>	<i>leña</i>	<i>bosque</i>	<i>selva</i>

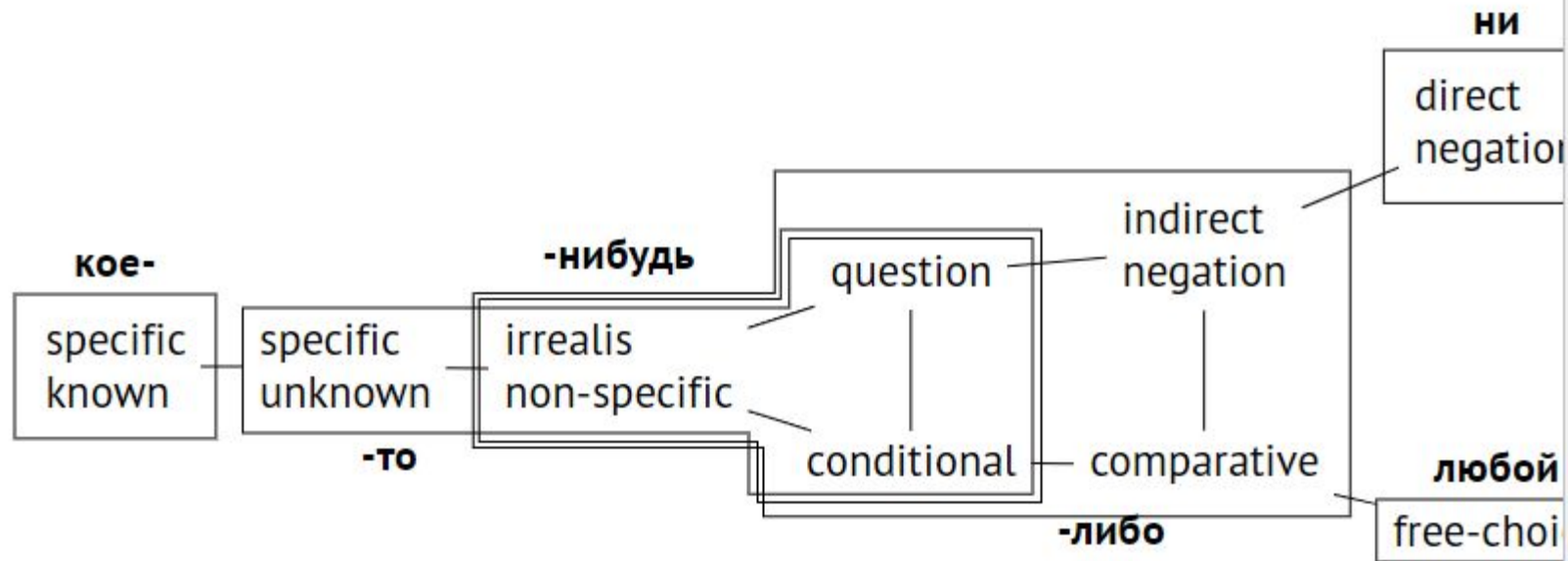
Неопределённые местоимения

Английский



Неопределённые местоимения

Русский



Достоинства семантических карт

- Несомненные достоинства:
 - ▶ позволяют проводить межъязыковое сопоставление «моделей полисемии» (в отличие от «инвариантного» подхода)
 - ▶ делают предсказания о возможной и невозможной дистрибуции языковых единиц
 - ▶ позволяют формулировать имплицативные зависимости
 - ▶ позволяют представить путь диахронического развития
- Часто говорят, что семантические карты отражают некие универсально значимые когнитивные отношения или описывают «когнитивные пространства» (conceptual space)
- На каком-то абстрактном уровне это, скорее всего, верно, но совершенно не очевидно, что семантические карты отражают какую-либо непосредственную «ментальную реальность»
- С тем же успехом они могут отражать результаты частотных диахронических процессов

Метод MDS

Метод MDS позволяет обобщить сходства мультязычных параллельных контекстов и сгруппировать функционально и/или семантически близкие контексты на основе информации о грамматических формах и конструкциях, используемых для маркирования языкового явления, в лексической типологии — на основе лексических единиц. В отличие от традиционных семантических карт, карты, построенные таким способом, учитывают большой спектр различных ситуаций, а не ограниченный набор специально отобранных примеров.

Lexical typology through similarity semantics: Toward a semantic map of motion verbs*

В работе [Wälchli, Cysouw 2012] авторы используют корпус Библии для изучения глаголов движения в сотне языков. Мультиязычный параллельный корпус позволяет с помощью применения метода многомерного шкалирования построить семантическую карту, основанную на матрице сходства контекстов.

Построение матрицы сходства

Важным этапом при построении семантической карты является разметка данных и построение матрицы сходства. В работе [Wälchli, Cysouw 2012] описан процесс построения такой матрицы для эксперимента с глаголами действия. Каждая переводная единица имеет реализации ситуации на N языках.

Для каждой пары переводных единиц можно посчитать расстояние, например, количество пар с несовпадающим маркированием и поделить его на общее число пар. Таким образом, мы получим матрицу со значениями от 1 до 0:

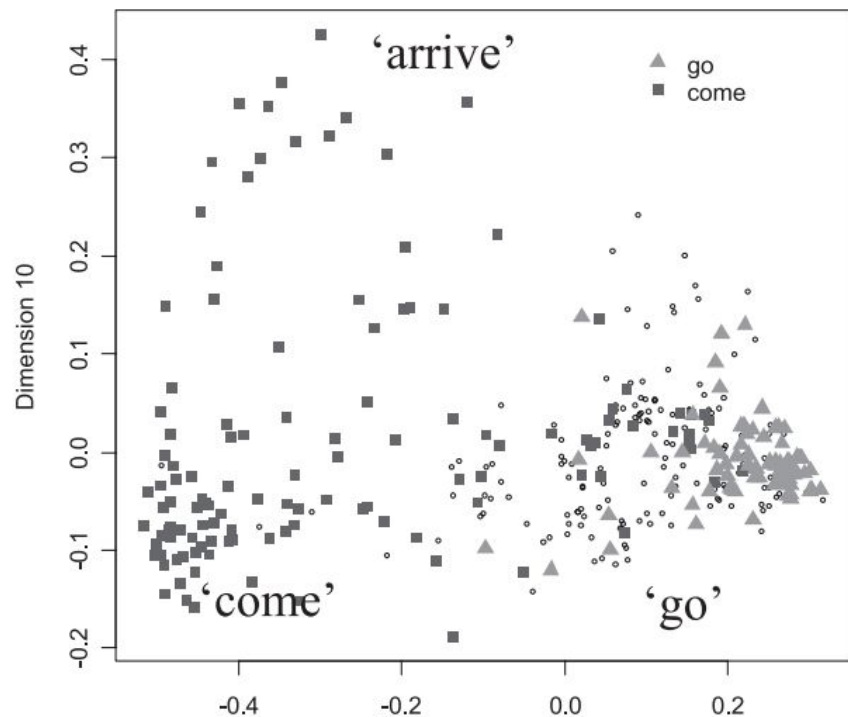
Пример данных [Wälchli, Cysouw 2012]

Situations	English	French	Hungarian	Mapudungu	Mari	Spanish (Senc.)	Spanish (RV)
1:31	come	s=approcher	megy	fülkon	mijaš	ir	llegar
4:4	come	venir	jön	aku	čongeštaš	venir	venir
5:1	come	arriver	ér	puw	mijaš	llegar	venir
6:29	come	venir	jön	amu	tolaš	ir	venir
6:48b	come	se=diriger	indul	NA	mijaš	ir	venir
9:13	come	venir	jön	küpa	tolaš	venir	venir

Матрица расстояний содержит $N*(N - 1)/2$ пар, где N — количество переводных единиц, если переводных единиц 500, получится 124750 пар, такой объем данных сложно обработать и упорядочить вручную без компьютерной обработки. Используя MDS можно обобщить имеющиеся данные и зафиксировать важную информацию о возможных противопоставлениях в нескольких измерениях.

Измерения матрицы MDS, отражают, как переводные единицы могут быть противопоставлены друг другу на основе разных принципов, каких — вопрос интерпретации результатов.

English (King James)



Spanish (Lenguaje Sencillo)

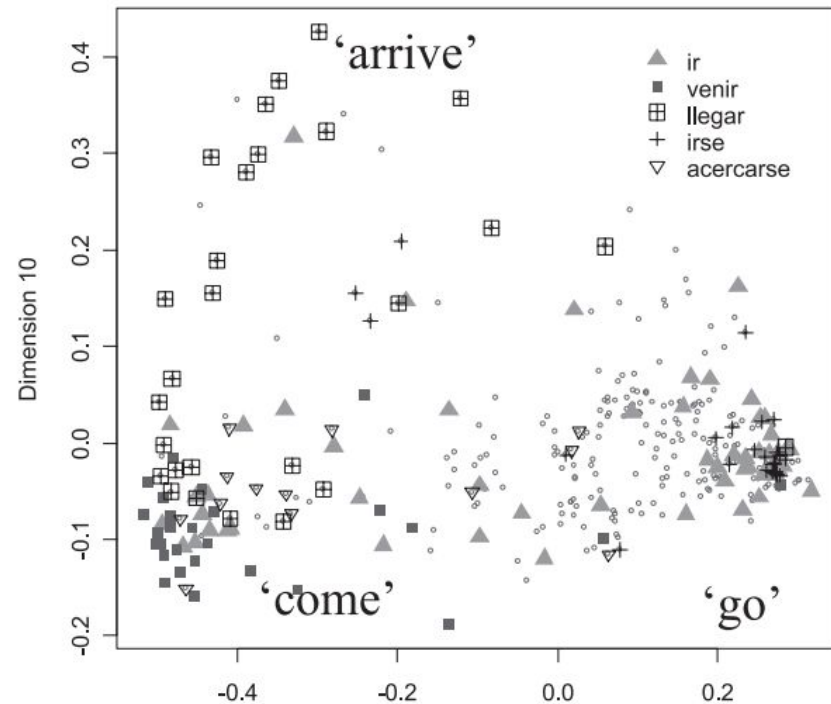


Table 4. *Interpretation of the first twelve dimensions*

	Negative Pole	Positive Pole		Negative Pole	Positive Pole
1	come/arrive	go/depart	7	walk/pass/cross	
2	transport (bring)		8	assemble	
3	enter		9	ascend/descend	
4	exit/depart		10	come	arrive
5	follow		11	walk	pass/cross
6	run/flee		12	carry	lead

Семантическая карта позволяет не только проследить языковую вариативность, но и распределить сравниваемые ситуации по семантически и/или функционально близким группам. Ситуации, которые систематически маркируются похожим образом в рамках переводной единицы, как правило, оказываются близкими по значению или по функциональному употреблению средств выражения. Установить такого рода закономерности вручную представляется сложным, а автоматически построенная семантическая карта позволяет быстро и содержательно упорядочить рассматриваемые данные.

Итоги

В конечном итоге семантические карты подтверждают верность фундаментального допущения, на котором основана современная грамматическая теория,— что существует универсальный набор смыслов, выражаемых грамматическими категориями естественных языков, который является структурированным и упорядоченным и об элементах которого можно говорить в терминах «более близкий» и «более далекий». О том, как организована эта система, многое уже известно, но намного больше еще только предстоит узнать.

[\[Татевосов 2002\]](#)