

# Построение датасета часть 2

# Подготовка данных

- Выбираем языки
- Выравнивание по предложениям есть, обрабатываем парсером UDPipe, для получения синтаксической и морфологической разметки.
- Опираясь на автоматическую разметку грамматических характеристик, извлекаем примеры с употреблениями интересующего нас явления.
- **Про ошибки:** Лишние нерелевантные примеры (false positive) придется отфильтровать вручную. Что-то может не попасть в нашу выборку при автоматическом поиске, с этим придется смириться.

## Этапы построения датасета

1. Смотрим, как соотносятся средства выражения исследуемого явления с тем, что имеем в разметке. И понимаем, что можем извлечь напрямую, а для чего придется написать дополнительные правила.
2. Извлекаем из корпуса примеры, имеем в виду, что мы можем захватить лишнего, а что-то можем не найти. В первом случае, можно будет перепроверить все, а во втором – не получится.
3. Считаем распределения конструкций по языкам, делаем первичные выводы и думаем над разметкой, которая нам понадобится.
4. Первое, что придется разметить – нетривиальные конструкции, которые попались в переводах в качестве соответствия вашей исходной искомой конструкции

# Modality

*Van der Auwera, Plungian, 1998*

Possibility			
Non-epistemic possibility			Epistemic possibility (Uncertainty)
Participant-internal possibility (Dynamic possibility, Ability, Capacity)	Participant-external possibility		
	(Non-deontic possibility)	Deontic possibility (Permission)	
Participant-internal necessity (Need)	(Non-deontic necessity)	Deontic necessity (Obligation)	Epistemic necessity (Probability)
	Participant-external necessity		
Non-epistemic necessity			
Necessity			

# Non-epistemic modality

- **Participant-internal**

- **Possibility (Dynamic possibility, capacity, ability)**

*Boris can get by with sleeping 5 hours a night.*

- **Necessity**

*Boris needs to sleep 10 hours every night for him to function properly.*

# Non-epistemic modality

- Participant-external, non-deontic (circumstances)
  - Possibility  
*To get to the station, you can take bus 6.*
  - Necessity  
*To get to the station, you have to take bus 6.*
- Participant-external, deontic (a person, ethical norms)
  - Possibility (Permission)  
*John may leave now.*
  - Necessity (Obligation)  
*John must leave now.*

# Epistemic modality

- **Epistemic possibility (Uncertainty)**

John may have arrived.

- **Epistemic necessity (Probability)**

John must have arrived.

# Possibility

- **Non-epistemic, Participant-internal**

*Boris can get by with sleeping 5 hours a night.*

- **Non-epistemic, Participant-external, non-deontic (circumstances)**

*To get to the station, you can take bus 6.*

- **Non-epistemic, Participant-external, deontic (a person, ethical norms)**

*John may leave now.*

- **Epistemic possibility (Uncertainty)**

John may have arrived.



# Necessity

- **Non-epistemic, Participant-internal**

*Boris needs to sleep 10 hours every night for him to function properly.*

- **Non-epistemic, Participant-external, non-deontic (circumstances)**

*To get to the station, you have to take bus 6.*

- **Non-epistemic, Participant-external, deontic (a person, ethical norms)**

*John must leave now.*

- **Epistemic necessity (Probability)**

*John must have arrived.*