Advantages of the flux-based interpretation of dependency length minimization

Sylvain KAHANE, Chunxiao YAN
MoDyCo, Université Paris Nanterre

Quasy, Syntaxfest, Paris, August 26, 2019

Outline

- Dependency length minimization (DLM)

- Cognitive relevancy of DLM

- DLM-related constraints

- Conclusion

Dependency length minimization (DLM)

Studies of dependency length minimization(DLM) in natural languages (Liu,2008 ; Futrell et al., 2015)

Properties correlated with DLM

- Much less non-projective structures in natural languages than in randomly ordered trees (Ferrer i Cancho, 2006 ; Liu, 2008)

- DLM is a factor affecting the grammar of languages and word order choices (Gildea & Temperley, 2010 ; Temperley & Gildea, 2018)

DLM and dependency flux

dependency flux between two words = set of dependencies that link a word on the left with a word on the right (Kahane et al., 2017).

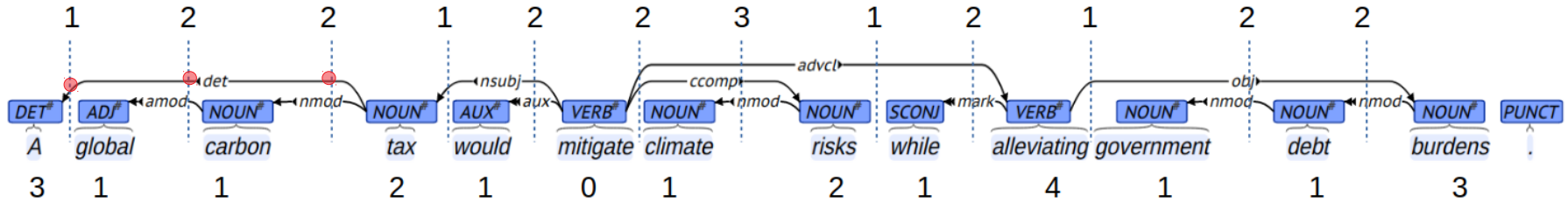flux size at position P = number of dependencies that cross P
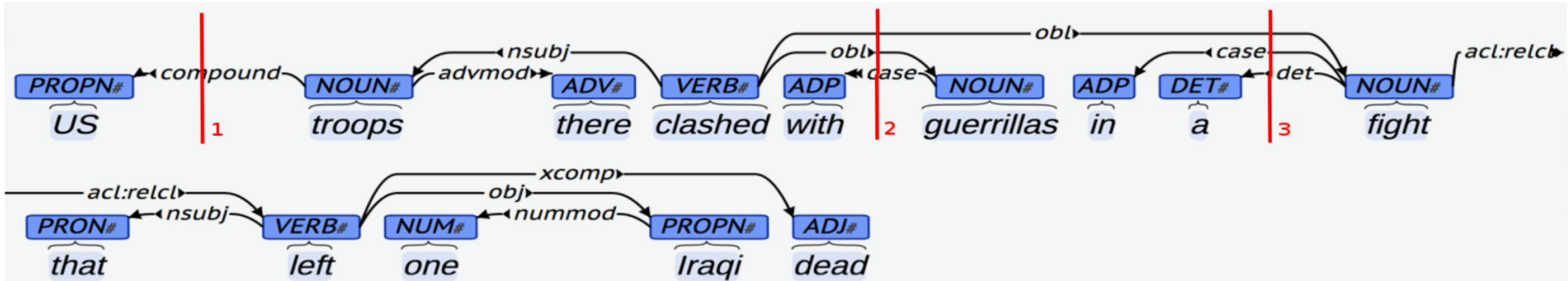
Position 1: flux size = 1   Position 2: flux size = 3   Position 3: flux size = 3
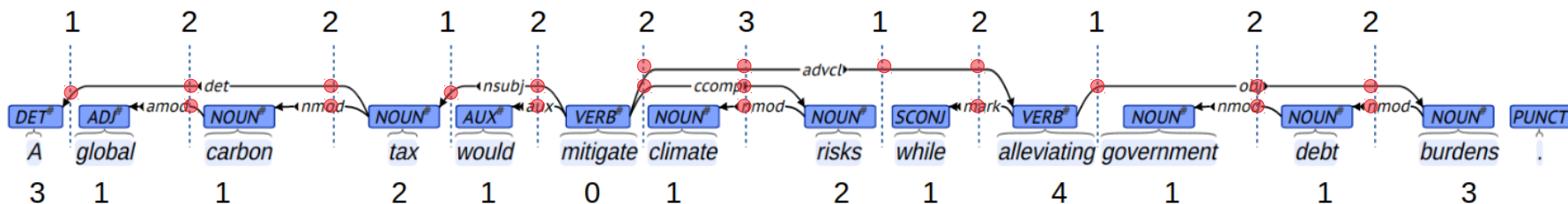
DLM and dependency flux

It is easy to check that the dependency length is always equal to the dependency flux size. How ?

Relation *det* : length =3, = cross 3 inter-word fluxes (red points)

DLM and dependency flux

It is easy to check that the dependency length is always equal to the dependency flux size.

Flux size of sentence = 1(*det*)+2(*det, amod*)+2 (*det, nmod*)+1(*nsubj*)+2(*nsbuj, aux*)+2(*advcl, ccomp*) +3(*advcl, ccomp, nmod*)+1(*advl*)+2(*advcl, mark*)+1(*obj*)+2(*obj,nmod*)+2(*obj, nmod*) = **21(red points)**

Dependency length of sentence = 3(*det*)+1(*amod*)+1(*nmod*)+2(*nsubj*)+1(*aux*)+0+1(*nmod*)+2(*ccomp*) +1(*mark*)+4(*advcl*)+1(*nmod*)+1(*nmod*)+3(*obj*) = **21(red points)**

Two different views on DLM.

Cognitive relevance of DLM

- **DLM ==> minimization of the flux size** of the sentence and therefore of all inter-word fluxes

- Frazier & Fodor (1978) : Sentences are more or less parsed as fast as they are received by the speakers.

- The flux in a given inter-word position is the information resulting from the portion of the sentence **already analyzed** that is necessary for its **further analysis**.

- Obvious link between the **flux** and the **working memory** of the recipient of an utterance (as well as the producer of the utterance).

Cognitive relevance of DLM

Limitations of working memory

- Miller (1956) observed that memory span of young adults is approximately **7** items.
- *A central memory store limited to **3 to 5** meaningful items in young adults.* Cowan(2001)

Cognitive relevance of DLM

Flux based interpretation :

Dependency flux in inter-word positions is a good approximation of what the recipient must remember to parse the rest of the sentence.
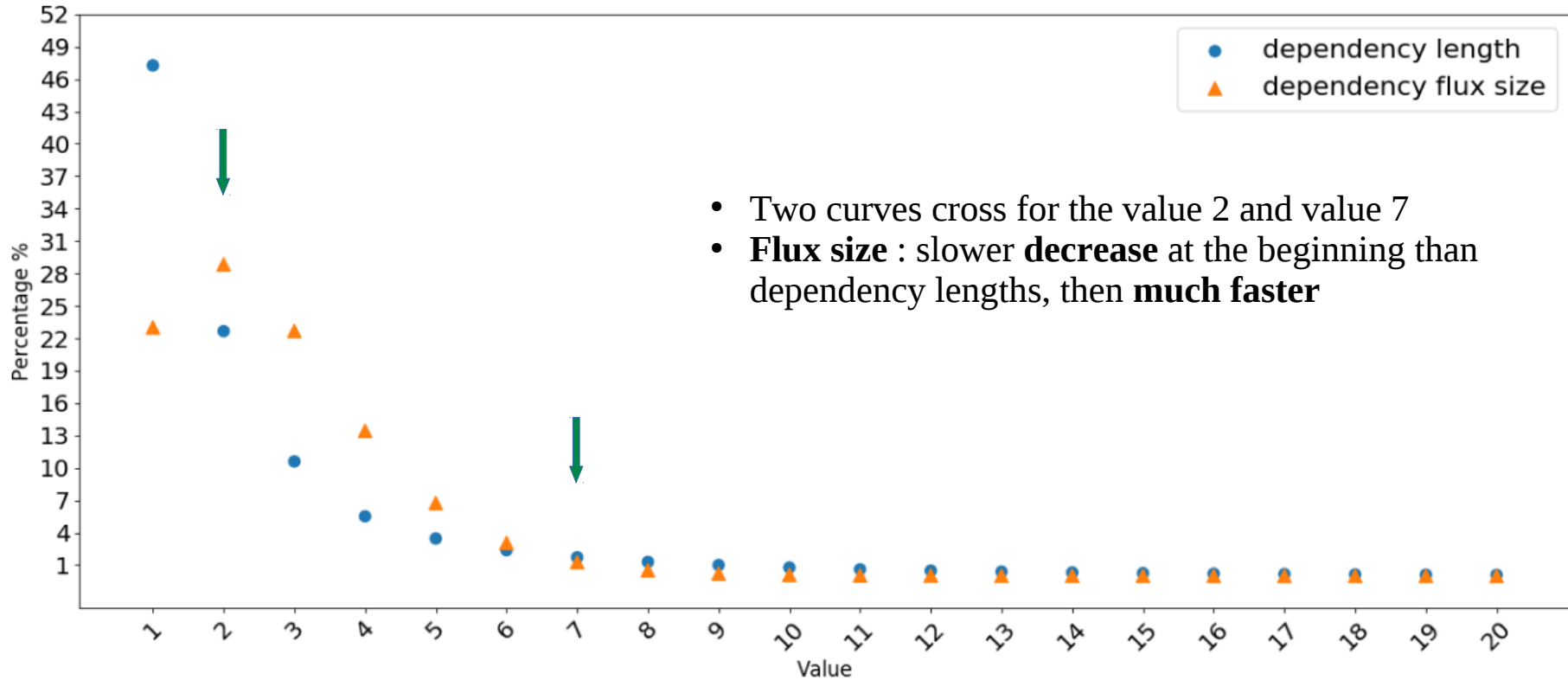
Dependency length based interpretation:

It is cognitively expensive to keep a dependency in working memory for a long time and that the longer a dependency is, the more likely it is to deteriorate in working memory (Gibson, 1998; 2000).

DLM-related constraints

- **<u>Constraints on size of inter-word fluxes</u>**

- **Constraints on center-embedding and constrains on structure fluxes**

- **Constraints on the potential flux**

Distribution in all UD data

- Two curves cross for the value 2 and value 7
- **Flux size** : slower **decrease** at the beginning than dependency lengths, then **much faster**

Flux size and dependency length

In all UD data:

- 99% of flux sizes ≤ **7**

- 99 % of dependency lengths ≤ **17**

Flux size and dependency length

Similar results in the 47 UD treebanks containing more than 100,000 flux positions:

• Two curves cross for the value **2** , and second croissing between **5** (UD_Finish-FTB) and **8** (in 9 treebanks: *UD_Urdu-UDTB, UD_Persian-Seraji, UD_Hindi-HDTB, UD_German-HDT, UD_German-GSD, UD_Dutch-Alpino, UD_Chinese-GSD, UD_Arabic-PADT and UD_Japanese-BCCWJ*).

• Flux size : slower decrease at the beginning than dependency lengths, then much faster

• 99% dependency lengths $\leq n$, n between **9** (*UD_finish-FTB*) and **27** (*UD_Arabic-PADT*).

• 99% flux sizes $\leq n$, n between **6** (12 treebanks) and **11** (*UD_Japanese-BCCW*).

# Flux size and dependency length

If DLM expresses a constraint on the average value of dependency lengths and flux sizes, we see that there is also a fairly **strong constraint** on the **size of each flux**, whereas there is not such a strong constraint on the length of each dependency.

For this reason, we postulate that DLM results more on a constraint on flux sizes than on dependency lengths, even if it is not possible to give a precise limit to the size of individual fluxes as Kahane et al. (2017) have already shown.

DLM-related constraints

- **Constraints on size of inter-word fluxes**

- **Constraints on structure fluxes**

- **Constraints on the potential flux**

Center-embedding constraints
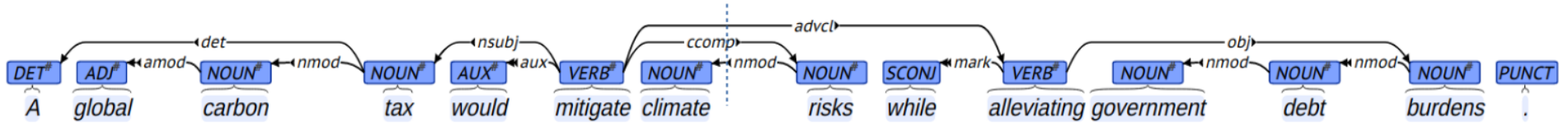
Center-embedding construction in terms of flux

Disjoint dependencies : **no common vertex**

	risks	*alleviating*
climate	**<nmod**	
mitigate	>ccomp	**>advcl**

The number of disjoint dependencies in a flux is very constrained (Kahane et al., 2017): 99.62% of the fluxes in the UD database have **less than 3 disjoint dependencies.**

DLM-related constraints

- **Constraints on size of inter-word fluxes**

- **Constraints on center-embedding and constrains on structure fluxes**

- **Constraints on the potential flux**

Potential flux

We do not know which **word already processed** will be **linked with a word not yet processed.**

Keeping all the words already processed and still accessible in the working memory (cf. principles of transition-based parsing ; Nivre, 2003)
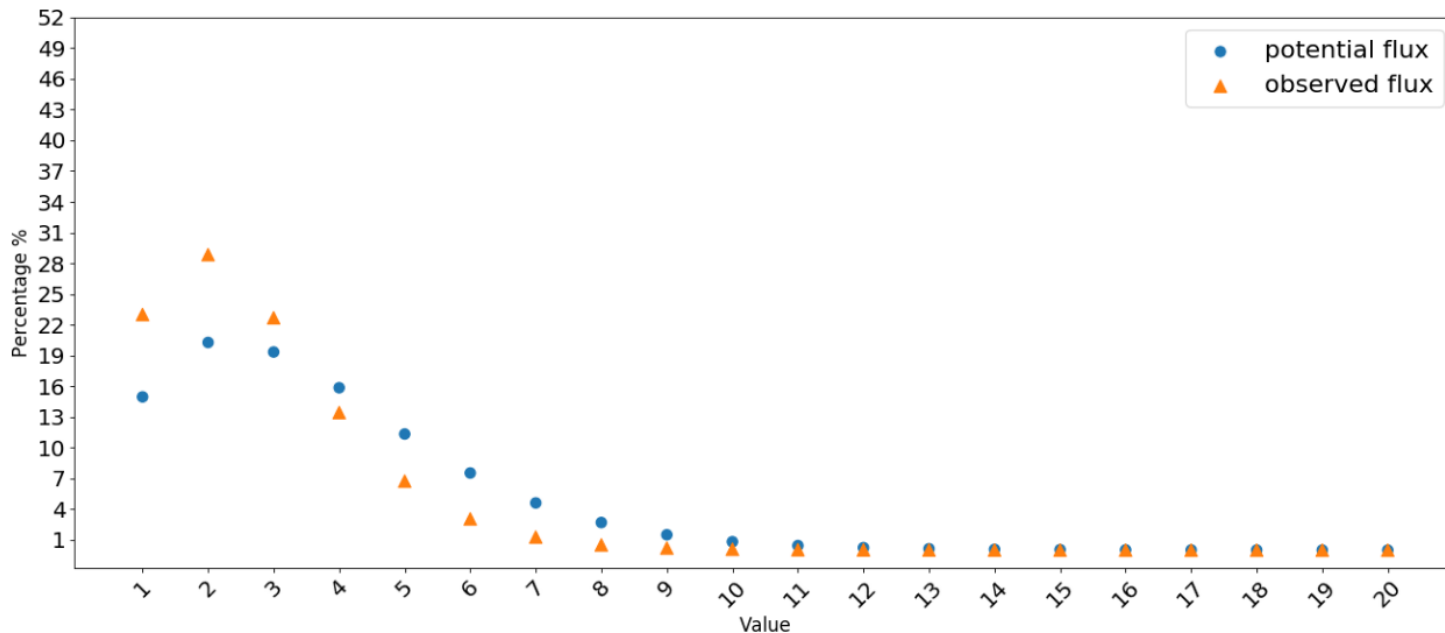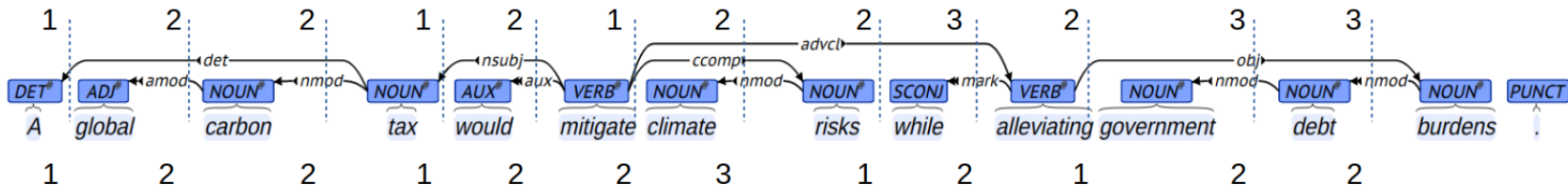
(Projective) potential flux : the set of words accessible while maintaining the **projectivity** of the analysis.
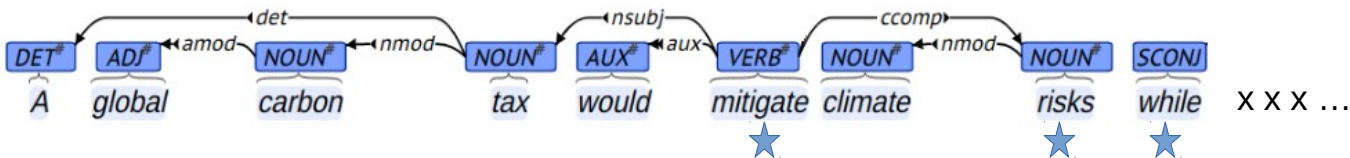
Potential flux at « while » : 3

# Potential flux and observed flux (all UD data)

Potential flux :

| 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 3 | 2 | 3 | 3 |

DET# ADJ# NOUN# NOUN# AUX# VERB# NOUN# NOUN# SCONJ VERB# NOUN# NOUN# NOUN# PUNCT

A global carbon tax would mitigate climate risks while alleviating government debt burdens .

Observed flux (flux size):

| 1 | 2 | 2 | 1 | 2 | 2 | 3 | 1 | 2 | 1 | 2 | 2 |

flatter than observed flux

⇒ **Projective potential fluxes generally greater than observed flux**.

Potential flux : head-initial and head-final languages

Head-initial : Arabic, Irish ⇒ percentage increase slowly at beginning, and then decrease slowly ⇒ greater values than head-final

Head-final : Jepanese, German ⇒ similar to general distribution of entire UD
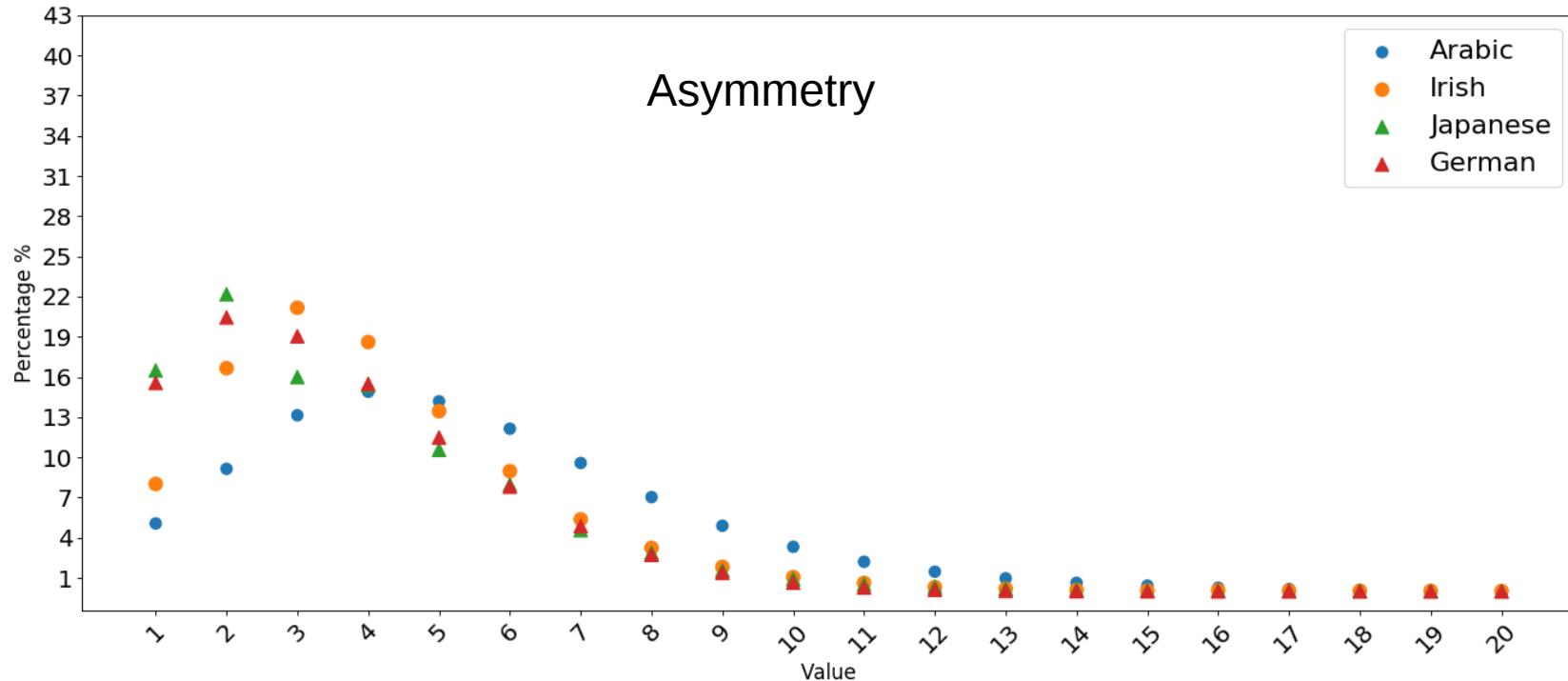
Conclusion

Dependency length minimization (DLM) is also a property of inter-word dependency fluxes.

An asymmetry between head-initial and head-final languages concerning the flux that could be related to the different potential flux in these two kinds of languages.

We believe that the constraints on the flux are far to be limited to its average size and that the structure of the flux plays an important role in its complexity.

Thanks !