German Support Verb Constructions in Context Embeddings

Xinyao Lu [1]

[1]Bachelor Student (Computational Linguistics + German Studies) at FAU-Erlangen-Nuremberg

Friedrich-Alexander-Universität
Philosophische Fakultät und
Fachbereich Theologie

Querying Support Verb Constructions

1. A rough **semantic definition** for human beings
   - Support verb construction (**SVC**, German: *Funktionsverbgefüge*) is a special kind of **verbo-nominal collocation**, in which the meaning of the verb is almost completely transferred to the noun and itself appears only as a support verb.
   - Typical examples of SVCs in German:
     zum Ausdruck bringen, in Kauf nehmen, in Rechnung stellen
2. A strict **syntactic definition** for corpus querying
   - A SVC consists of at most five components: a preposition (**APPR**), an article (**ART**), a verbal noun (**NN**), a prepositional complement, (**PP-COM**) which modifies the NN, and a support verb (**V**).
   - The ART and the PP-COM are optional for a SVC, the other three components (APPR, NN and V) are compulsory. The APPR must build a prepositional phrase (**PP**) with the ART and NN. This PP is the prepositional object (**OP**) of the V. Sometimes the APPR and the ART can be abbreviated in one token (**APPRART**).

Concordance Reading and Syntax Analysis

Query: ['stellen', 'gestellt', 'stellt', 'stellte']

[0] 🔎 subset (1130):

[1] 🔀 arrangement : ⚙ Random Sort {'seed': 42}

[2] 🔎 subset (8): ⚙ Select by a Token-Level Attribute {'offset': -1, 'value': 'Frage', 'tokens_attribute': 'word', 'case_sensitive': False, 'regex': False, 'comparison_operator': '==', 'negative': False}

Line ID	Left Context	Node	Right Context
55	erführen zu sollen. Aber ehe die Tante noch eine direkte Frage	**stellen**	konnte, wurde der Rentmeister gemeldet, der ihr in diesem Au
546	vas jene gepflanzt, hatte diese entweder zerstört oder in Frage	**gestellt**	. Jeder hohe Grad von Civilisation bringt eine solche Complica
92	Rätsel war dieses Mädchen. Schon wollte Forbeck eine Frage	**stellen**	, da ließen sich vor der Klause hastige Schritte vernehmen. Es
406	eles, ja alles, so ahnte Anton, ward durch diesen Krieg in Frage	**gestellt**	. Deshalb stürzte er seinem Prinzipal entgegen, als dieser die T
787	weigend die Achseln, als mein trostloser Blick ihm diese Frage	**stellte**	. Wir richteten nun die Kranke in meinem Dachzimmer ein, um

Figure 1. Reading Concordance Lines in FlexiConc

	Grammatical Role	Construction	SVC	Matches	Procentage
Figure 2	Frage in SVC	OA in Frage stellen	True	107	22,81%
Figure 3	Frage as Subject	es stellt sich die Frage	False	118	25,16%
Figure 4	Frage as Direct Object	Frage stellen	False	242	51,60%
	Syntax Error (?)	jmd stellt sich (in) der Frage	False	2	0,00%

Table 1. "Frage < < s > > {stellen/V}" in *Tageszeitung* Corpus, year = 2011

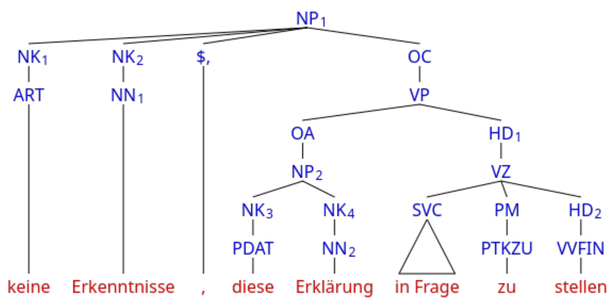

Figure 2. *Frage* in a SVC dependent on the support verb

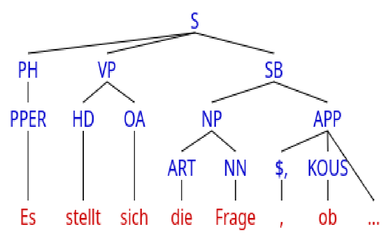

Figure 3. *Frage* is subject of a reflexive verb

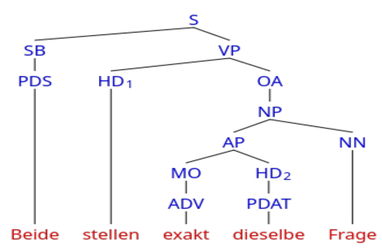

Figure 4. *Frage* is direct object of a full verb

See more in Repository and Brochure

- Scan the QR code on the left to star the project repository and follow further works.
- Scan the QR code on the right to get the brochure for more information, detailed statistics and the references.

Organising Dataset: Assumptions and Challenges

1. Assumption 1: Dataset should contain three components: a **control** group, a **comparison** group and a **contrast** group (more details see brochure)
2. Assumption 2: Most SVCs can be found as the **collocation** of a support verb and a verbal noun.
3. Experiment to Assumption 2: Search for SVCs in a collocation list with the example {stellen/V}
   - Step 1: Download a list of collocations from **CQPweb**
     Tageszeitung corpus (1986 - 2011), **lemma query** {stellen/V}, **Window size**: 5 tokens on both sides, **POS tag** of the collocates: NN (common noun), **minimal frequency threshold**: 5; list sorted by **Observed Collocate Frequency**.
   - Step2: Compare the collocations with a table of **familiar** SVCs
     The *TIGER-Annotationsschema* provides in its appendix a table of familiar SVCs. All the verbal nouns in this table, which build SVCs with *stellen*, can be found in the collocation list from the last step.
   - Step 3: Analyse the collocation list with **Association Measures**

     Table 2 includes the first 5 entries in the collocation table, which also appear as verbal nouns in the table of familiar SVCs. **No.** shows the ranking of observed collocate frequency in the original table.

     Mutual Information (MI)

$$MI = \log_2 (Observed\ Frequency\ /\ Expected\ Frequency)$$

     is chosen to reflect the degree of association between the NN and the support verb. (the complete table see brochure)

NN	No.	Expected	Observed	MI
Frage	14	838,22	19349	4,53
Verfügung	23	219,32	13320	5,92
Aussicht	78	67,22	3424	5,67
Beweis	103	68,82	2420	5,14
Wahl	174	332,49	1335	2,01

Table 2. Observed SVCs in the collocation list {stellen/V}

   - Step 4: Compare the results of collocation list with **proximity query**
     For known SVCs, proximity query (PP < <s> > {stellen/V}, e. g. in Rechnung «s» {stellen/V}) should be a reliable way to ascertain their frequency in the corpus. In table 3, the **observed collocate frequency** from the collocation list is called $\mathbf{O_1}$, while the number of **matches from the proximity query** is called $\mathbf{O_2}$; the value O_2 / O_1 should indicate how often the collocation is actually used in a SVC. (the complete table see brochure)

NN in collocation list of {stellen/V}, window size = 5			PP < < s > > {stellen/V} (in the same sentence)		
NN	Observed O_1	MI	Preposition	Matches O_2	O_2 / O_1
Aussicht	3424	*5,67*	in	4500	*1,31*
Verfügung	13320	*5,92*	zur	15957	*1,20*
Abrede	282	*7,25*	in	326	*1,16*
Beweis	2420	*5,14*	unter	2794	*1,15*
Rechnung	1002	*3,77*	in	1066	*1,06*

Table 3. Collocation list vs. Proximity Query

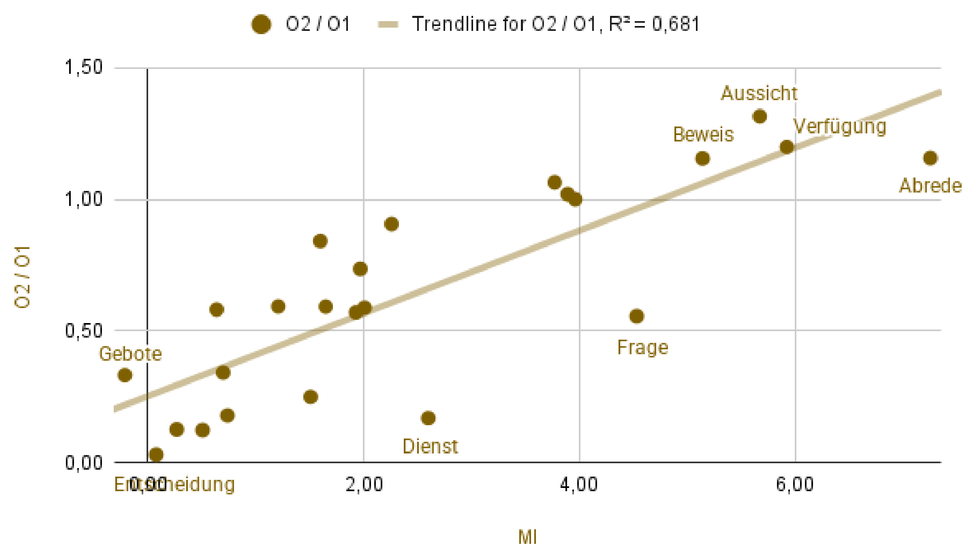

Figure 5. Linear Regression of MI and O_2 / O_1

4. Discussion: Combining Association Score and **Concordance Reading** to gather the dataset
   - From figure 5 we can draw a rule of thumb: The stronger the association of a verb and a noun, the more possible they will build up a SVC together.
   - Since high association score is (practically) a **necessary but not sufficient** condition of SVC, to find out the true SVCs in a collocation list, a manual concordance reading is still necessary.
   - We should pay extra attention to the collocates **far under the trend line** (*Entscheidung*, *Dienst*, *Frage* etc.) Their position under the trend line means that even though they have a high association score, they are not often (always) used in SVC.

xinyao.lu@fau.de