

(Pseudo-)VP-Topikalisierung im Polnischen und Tschechischen: Korpusevidenz und Grammatik

09. Juli 2004

- (1) [Bez względu na to, jaki był cel podjętej wyprawy,] prowadzić
ohne Rücksicht auf das welches war Ziel unternommen_{gen} Expedition_{gen} führen
to będzie do przedłużania bałkańskiego dramatu.
das wird zu Verlängerung Balkan_{gen} Drama
'[Abgesehen davon, was das Ziel der unternommenen Expedition war,] führen wird das
zur Verlängerung des Balkandramas.'

1 VP-Voranstellung: Einführung

Totale VP-Topikalisierung:

- (2) a. Sie hat nicht [_{VP} den Peter geküsst]
b. [_{VP} den Peter geküsst]_j hat sie nicht t_j

Partielle VP-Topikalisierung:

- (3) a. Sie hat nicht [_{VP} den Peter geküsst]
b. Sie hat [den Peter]_i nicht [_{VP} t_i geküsst]
c. [_{VP} t_i Geküsst]_j hat sie [den Peter]_i nicht t_j (Bsp. nach Fanselow 2004)

- Vorausgesetzt für die in (3) angedeuteten Bewegungen:

- In betrachteter Sprache existieren unabhängig Scrambling (vgl. (3-b)) oder auch Extraposition (Müller, 1998)
- Es existiert auch *totale* VP-Topikalisierung (vgl. (2-b))

- Fraglich: Liegt in Beispielen wie (3-c) evtl. Bewegung einer kleineren Einheit als VP (z.B. V⁰) vor?

Relevante empirische Eigenschaften (vgl. Müller 1998, Fanselow 2004):

- VP ist (nach Bewegung) selbst Insel für Extraktion
- Nur wenn das Vorfeld die Zielposition der VP ist, kann *partielle* VP-Voranstellung stattfinden. Vgl. dagegen Mittelfeldspitze (4-a,b) oder Linksversetzungsposition (5-a,b):

- (4) a. dass [den Peter geküsst] keiner hatte
 b. ?-*dass geküsst keiner den Peter hat
- (5) a. ein Buch schenken, das würde ich der Claudia nie
 b. ?-*schenken, das würde ich der Claudia das Buch nie

2 Tschechische und polnische Satzstruktur

2.1 Tschechisch

Topologie:

XP?	<i>že</i> 'dass'	XP?	aux. > refl. > pron. Klitika	XP*	Negation	Futuraux.	XP*	VP
-----	---------------------	-----	---------------------------------	-----	----------	-----------	-----	----

Relevante Kontraste:

- generell nur eine Konstituente oder *l*-Partizip vor Klitik-Cluster möglich
- in *wh*-Fragen und eingebetteten Sätzen: 1 Konstituente *unter* C⁰ möglich, aber nicht *l*-Partizip-Anhebung vor den Klitika-Cluster
- *l*-Partizip-Anhebung "Last Resort" zur Stützung der Auxiliarklitika
- unterschiedliche verbale Positionen durch die möglichen relativen Abfolgen mit Negation bestimmt

Hier relevante Konstruktionen:

Angehobene infinite VPn

1. (inkl. Negation) vor Perfekt-/Konjunktivauxiliar – (*hohe VP-Anhebung*)
2. nach Klitika-Cluster, aber vor (Negation und) Futurauxiliar (*tiefe VP-Anhebung*)

2.2 Polnisch

Topologie:

<i>że</i> 'dass'	XP*	Perfekt-/Konjunktivaux.?	pron. Klitika	XP*	Negation	Futuraux.	XP*	VP (+ Perfekt-/Konjunktivaux.)?
---------------------	-----	--------------------------	---------------	-----	----------	-----------	-----	---------------------------------

Relevante Kontraste:

- wie im CZ werden unterschiedliche verbale Positionen durch die möglichen relativen Abfolgen mit Negation bestimmt, vgl. (6)
- Perfektauxiliar kann alternativ in Agr_S⁰ oder in V⁰ generiert werden; keine Verbanhebung in (6-b)
- *by* generiert in T⁰-Kopf; (optional overte) Verbanhebung nach T⁰ in (6-d)

- Futurauxiliar in v^0 , Modalverben in V^0 (nach Błaszczak 2001)

- (6) a. My-śmy go (nie) widzieli wczoraj.
 wir-AUX_{praet} ihn nicht sahen gestern
 ‘Wir haben ihn gestern nicht gesehen.’
 b. My go (nie) widzieli-śmy wczoraj.
 c. Jan by go (nie) spotkał jutro.
 Jan AUX_{konj} ihn nicht sah morgen
 ‘Jan würde ihn morgen nicht sehen.’
 d. *Jan go (nie) spotkał-by jutro. (Błaszczak, 2001, 80)

Relevante Konstruktionen:

Angehobene infinite VPn

1. vor (Negation und) Futurauxiliar
2. vor Perfekt-/Konjunktivauxiliar (plus Negation)

abzugrenzen von Anhebung von *l*-Partizipien (vgl. (7)) – aber hier (anders als im CZ) auch Voranstellung einer Konstituyente *plus l*-Partizip möglich (vgl. (8)):

- (7) Spotkałby go jutro.
 treffen-AUX_{konj} ihn morgen
 ‘Er würde ihn morgen treffen.’
 (8) Co radziłbyś bliskiemu sobie młodemu człowiekowi, aby zrobił po
 was raten-AUX_{konj}-AUX_{2.sg} nahem REFL jungem Menschen dass tat nach
 ukończeniu szkoły?
 Ende Schule_{gen}
 ‘Was würdest Du einem Dir nahestehenden jungen Menschen raten, dass er nach dem
 Ende der Schulzeit tun soll?’

Aber nicht in anderer Linearisierung:

- (9) *Radził Pawłowi bym ...
 geraten Paweł-dat AUX_{konj}-AUX_{1.sg}

[pos = praet] [] "by"[pos = aglt] within s → 1 Treffer:

- (10) ..., pochował ja bym go tak, żeby go i na sąd ostateczny
 versteckt ich AUX_{konj}-AUX_{1.sg} ihn so dass ihn auch auf jüngste Gericht
 nie znaleźli.
 nicht finden
 ‘Verstecken würde ich ihn so, dass man ihn auch zum jüngsten Gericht nicht findet.’
 (Sienkiewicz 1895)

⇒ Partizip-Voranstellung ist auch im PL nicht totale VP-Anhebung

3 Ergebnisse der Korpusauswertung

3.1 Tschechisch

3.1.1 Hohe VP-Voranstellung

- Landeposition unter C⁰:

```
[tag = "Vf."] [tag != "Z.*"& word != "a"]*
"((js[iemt] [me]?)|(sis)|(ses))" within s
```

- (11) Samozřejmě uznávám, že ohánět jsem se po něm neměl.
 natürlich anerkenne_{1.sg} dass schlagen AUX-cl REFL-cl um ihn nicht-sollte
 ‘Natürlich erkenne ich an, dass schlagen ich mich um ihn nicht hätte sollen.’

- Totale VP-Voranstellung:

- (12) [zahodit ji a vydat se pěšky na útěk k východním
 wegwerfen sie und aufmachen REFL zu-Fuß auf Flucht zu Ost-
 hranicím směr domů]_k jsi nemohl _{t_k}
 Grenze Richtung nachhause AUX-cl nicht-konnte
 ‘Sie wegwerfen und sich zu Fuß auf die Flucht zur Ostgrenze Richtung Heimat
 machen konntest Du nicht.’

- Partielle VP-Voranstellung:

- (13) [_{VP} Usadit _{t_i} nastálo _{t_j}]_k jsem se_i chtěl _{t_k} [v Patagonii]_j
 setzen dauerhaft AUX-cl REFL wollte in Patagonien
 ‘Dauerhaft zur Ruhe setzen wollte ich mich in Patagonien.’

- Linearitätsbeschränkung: Voranstellung vor Klitika *immer* verbinitial:

P-Filter: [word = "[A-Z].*"] {1,6} [tag = "Vf.*"]

- (14) [_{VP} Pomoci objasnit celý případ] by mohli taxikáři, kteří
 helfen erklären ganze Fall AUX-cl konnten Taxifahrer welche
 napadení turistů viděli.
 Überfall Touristen_{gen.pl} sahen
 ‘Helfen, den ganzen Fall aufzuklären, könnten die Taxifahrer, die den Überfall
 auf die Touristen gesehen haben.’

Ergebnis: Ausschließlich Folgen zweier Infinitive – wie in (14); von Muttersprachlerin bestätigt:

- (15) *celý případ objasnit by mohli taxikáři ...
 ganzen Fall erklären AUX-cl konnten Taxifahrer

- Informationsstrukturelle Funktion:

Durchgängig I-Topik plus satzfinaler Fokus:

- (16) [O kolej žádá asi šestnáct set lidí.] [I-TOP Nabídnout]
 um Wohnheim bittet wohl sechzehn hundert Leute anbieten
 jsme mohli pouze [FOC pět set sedmdesát míst, která uvolnili
 AUX-cl konnten nur fünf hundert siebzig Plätze welche freimachten
 letošní absolventi.]
 diesjährige Absolventen
 ‘[Einen Wohnheimplatz beantragen ca. 1600 Leute.] Anbieten konnten wir nur
 570 Plätze, [welche die diesjährigen Absolventen freimachten.]’

3.1.2 Tiefe VP-Voranstellung

"((js[iemt][me]?)|(sis)|(ses))" [tag != "Z.*"& lemma != "a"]* [tag = "Vf.*"]
 [tag != "Z.*"& lemma != "a"]+ [tag = "Vp."] within s

- Totale VP-Voranstellung:

- (17) [...pak je to nejen proto,] že jsem [VP udělat kariéru] chtěla
 dann ist das nicht-nur darum dass AUX-cl machen Karriere wollte
 a chci, ale také proto, že ...
 und will aber auch darum dass
 ‘[...dann ist das nicht nur deshalb,] weil ich Karriere machen wollte und will,
 [sondern auch darum, weil ...]’

- VP-Voranstellung plus Linksbewegung von Argumenten

Klitika-Klettern & Topikalisierung:

- (18) Zadarmo_k jsem se_i jí_j [vzdát t_i t_j t_k]_l ale nechtěla t_l.
 umsonst AUX-cl REFL-cl PRON-cl ergeben aber nicht-wollte
 ‘Umsonst wollte ich aber nicht auf sie verzichten.’

Klitika-Klettern & Scrambling:

- (19) ...ale dál jsem se_i politice_j [věnovat t_i t_j] skutečně nechtěl.
 aber weiter AUX-cl REFL-cl Politik_{dat} widmen wirklich nicht-wollte
 ‘... aber weiter wollte ich mich der Politik wirklich nicht widmen.’

⇒ keine feste Linearisierung wie bei “hoher” VP-Voranstellung

- Entsprechendes gilt auch beim Futurauxiliar:

- (20) [...dole na kusu trávníku,] na kterém se hroby kopat teprve
 unten auf Stück Rasen auf welchem REFL-cl Gräber graben erst
 budou, ...
 werden
 ‘... unten auf dem Stück Rasen, auf welchem Gräber erst noch gegraben werden’

- Landeposition unterhalb des Subjekts möglich:

(21) Pokud by zákonodárci schválit misi nestihli, ...
 falls AUX-cl Gesetzgeber loben Mission nicht-schafften
 ‘Falls die Gesetzgeber es nicht schaffen würden, die Mission abzusegnen, ...’

- hier informationsstrukturelle Funktion *immer* minimaler Fokus auf dem rechts verbleibenden V^0 – 106 Belege, 0 Gegenbsp. → VP-Entleerung, Rechtsbewegung des finiten Verbs auszuschließen. Dabei V^0 meist negiert, aber auch

(22) “...Za druhou půli jsme však [VP vyhrát]_i určitě [FOC zasloužili t_i
 während zweite Hälfte AUX-cl aber gewinnen sicher verdient
],” řekl trenér
 sagte Trainer
 ‘“In der zweiten Hälfte haben wir aber zu siegen sicherlich verdient”, sagte der Trainer’

3.2 Polnisch

3.2.1 Hohe VP-Voranstellung

```
[pos = inf] [pos != interp & base != "(i)|(lub)|(albo)"]* "by"
[pos = aglt] within s
```

- Mit totaler VP-Voranstellung:

(23) Przecież inaczej w ogóle żyć byśmy nie mogli.
 doch anders in Allgemeinen leben AUX_{konj}-AUX_{1.pl} nicht konnten
 ‘Anders hätten wir doch gar nicht leben können.’

(24) W każdym bądź razie bodaj na części zasobów muzealnych uwłaszczyć
 in jedem möglichen Mal wohl auf Teil Vorräte museale bemächtigen
 byśmy się mogli, ...
 AUX_{konj}-AUX_{1.pl} REFL-cl konnten
 ‘Aber jedes beliebige Mal hätten wir uns wohl einen Teil der Museumsbestände aneignen können’

- Mit partieller VP-Voranstellung:

(25) Żyć bym bez nich nie potrafił.
 leben AUX_{konj}-AUX_{1.sg} ohne sie_{pl} nicht schaffte
 ‘Ohne sie zu leben, hätte ich nicht geschafft.’

(26) ..., ale ja niczego absolutnie wykluczać bym w tej materii
 aber ich nichts absolut ausschließen AUX_{konj}-AUX_{1.sg} in der Materie
 nie chciał.
 nicht wollte
 ‘... aber ich würde absolut nichts in dieser Materie ausschließen wollen’

⇒ keine Linearitätsbeschränkung wie im CZ!

3.2.2 Tiefe VP-Voranstellung

"by" [pos = aglt] [pos != interp & base != "((i)|(lub)|(albo))"]*
 [pos = inf] [pos != interp & base != "((i)|(lub)|(albo))"]*
 [pos = praet] within s

- Zielposition unter Subjekt möglich:

(27) Spółka koordynować będzie działalność pozostałych spółek
 Gesellschaft koordinieren wird Tätigkeit übriger Gesellschaften
 internetowych grupy Prokom Software [...] *Gen*
 Internet-Gruppe
 ‘Die Gesellschaft wird die Tätigkeit der übrigen Internet-Clubs der Gruppe Prokom Software koordinieren’

- und natürlich unter hoher AUX-Position:

(28) ... , ale też długo bym leżeć nie chciała, [bo
 aber auch lange AUX_{konj}-AUX_{1.sg} liegen nicht wollte weil
 bym chyba nie wytrzymała.]
 AUX_{konj}-AUX_{1.sg} wohl nicht aushielt
 ‘...aber lange würde ich auch nicht liegen wollen, [weil ich es wahrscheinlich nicht aushalten würde.] ’

3.2.3 Informationsstrukturelle Funktion

- Bei “hoher” VP-Voranstellung:

→ Gesamte vorangestellte VP I-Topik:

(29) ... bez sieci nie potrafię już żyć, a pewnie [ITOP umrzeć] [FOC
 ohne Netzwerk nicht kann schon leben und sicherlich sterben
 też] bym nie potrafił.
 auch AUX_{konj}-AUX_{1.sg} nicht könnte
 ‘... ohne Netzwerk kann ich nicht mehr leben, aber sterben könnte ich sicherlich auch nicht.’

→ Oder Teil der vorangestellten VP I-Topik:

(30) [Ponieważ jednak los tej ustawy w chwili obecnej jest wysoce
 da aber Los dieses Gesetzes in Zeit gegenwärtig ist hoch
 hipotetyczny,] dalej w tej kwestii posunąć bym się
 hypothetisch weiter in dieser Frage vorschieben AUX_{konj}-AUX_{1.sg} REFL-cl
 nie mógł.
 nicht konnte

‘[Da aber das Schicksal dieses Gesetzes zur Zeit noch hoch hypothetisch ist,] könnte ich mich weiter in dieser Angelegenheit nicht vorwagen.’

- bei tiefer VP-Voranstellung:

→ satzfinale V^0 oder postverbales Argument oft minimaler Fokus; gescrambelte VP Hintergrund

→ aber auch weite Foki:

- (31) [“Miasto Lozanna zresztą dość nudne. ...byłoby nam tu dobrze,] Stadt L. übrigens ziemlich langweilig wäre uns hier gut gdybyśmy przywyknąć mogli do cudzej ziemi!” wenn-AUX_{konj}-AUX_{2.pl} gewöhnen konnten zu fremdes Land ‘[“Die Stadt Lausanne ist übrigens recht öde. ... uns ginge es hier gut,] wenn wir uns gewöhnen könnten an das fremde Land.” ’

→ Oder Aufzählung:

- (32) [...leży i o Bożym świecie nie wie! ...] – I jadł kiełbasy nie liegt und über Gottes Welt nicht weiß und essen Wurst nicht będzie, i gorzalki nie posmakuje? wird und Schnaps nicht probiert ‘[liegt hier und weiß nichts von Gottes Erde ...] Weder die Wurst wird er essen noch den Schnaps wird er kosten?’

4 Schlussfolgerungen

- Syntaktisch:

- CZ: hohe VP-Voranstellung ist Bewegung einer Konstituente in eine Spezifiziererposition, verwandt der deutschen “VP-Topikalisierung”; Linearisierung entspricht der Basisgenerierung in der VP
- CZ: tiefe VP-Voranstellung kann als mehrfaches Scrambling gedeutet werden – keine feste Linearisierung
- PL: hohe wie tiefe VP-Voranstellung können als mehrfaches Scrambling gedeutet werden; nicht nur *eine* hohe Spezifiziererposition verfügbar

- Informationsstrukturell:

- Hohe VP-Voranstellung dient im CZ wie im PL der I-Topikalisierung (eines Teils) der VP
- Tiefe VP-Voranstellung ist im CZ Scrambling von Hintergrundmaterial aus dem Fokusbereich heraus; im PL scheinen stilistische Faktoren eine Rolle zu spielen, die fokusbegrenzende Funktion ist nicht obligatorisch

5 Ein Schritt zurück: Was haben wir benutzt?

5.1 Umfang und Aufbereitungsstand der Korpora

- reguläre Ausdrücke über Wortformen
- POS-Tags
 - Tschechisch:
 - * SYN2000, 100 Mio. Formen umfassendes Teilkorpus des ČNK
 - * Zusammensetzung: 60 % Publizistik, 15 % Belletristik, 25 % Fachliteratur
 - * POS-getaggt von Jan Hajič, positionelles Tagset mit ca. 2000 verschiedenen Tags; Korrektheit etwas über 93 % (Stand: 1998)
 - * derzeit regelbasierte Überarbeitung (Petkevič, Oliva) – vgl. Hajič *et al.* (2001)
 - Polnisch:
 - * frei verfügbares Korpus des Instituts für Grundlagen der Informatik, Akademie der Wissenschaften, Warschau (IPIPAN), vgl. Przepiórkowski (2004), ca. 70 Mio. Segmente (kleiner als orthographische Wörter), POS-getaggt
 - * Zusammensetzung: 20 % Prosa, 10 % wissenschaftliche Texte, 50 % Presse, 15 % Parlamentsstenogramme, 5 % Gesetze
 - * Besonderheit: Tags vor Disambiguierung erhalten und absuchbar
 - * eigener Tagger (L. Dębowski), keine Korrektheitsangabe
- Lemmatisierung
- Filterung der Suchergebnisse (positive oder negative Filter) – bisher nur in Gcqp
 - bisher nur in Gcqp, noch nicht in Poliqarp
- Baumdatenbank
 - Version 1.0 der Prague Dependency Treebank über LDC verfügbar, frei abfragbar über Oberfläche Netgraph (R. Ondruška, J. Mírovský)
 - Umfang: derzeit ca. 100.000 analysierte Sätze, handkorrigiert
 - Version 2.0 (soll noch 2004 zur Verf. stehen) mit überprüfter Annotation von Topik, Fokus und Kontrast (Guidelines dazu s. Hajičová *et al.* 2000)
 - voraussichtlich wenig geeignet für *empirische* Untersuchungen zur Informationsstruktur: 1. Kein Kontext abfragbar, 2. Annotationsrichtlinien theoriegebunden, vgl. “whatever depends on the verb from the left, has, as a rule, T; the value F remains only where it can be clearly recognized that ‘new information’ or a new relation is involved; the value C is assigned to the contrastive part of the topic (which expresses an element taken from a set of alternatives)”

5.2 Benutzerfreundlichkeit

- Konstruktion der Suchanfrage mit grafischem Tool in Gcqp
- Anfragesprache: cqp mit positionellem Tagset (Gcqp) vs. komfortable Aliase (Poliqarp):
[tag = "...P4.*"] (Gcqp)
[case = acc & num = p1] (Poliqarp)
komplizierte, ausdrucksstarke Anfragesprache in der PDT
- Komfortables Subkorpora-Management mit Gcqp
- Weiterentwicklung Bonito (P. Rychlý) mit Korpuszusammenstellung und verbesserten statistischen Funktionen

6 Desiderata

- Abfragekomfort, insbesondere beim Tagset – vgl. dazu etwa das neue Russische Nationalkorpus
- Subkorpus-Management
- Baumbanken mit Kontexten
- Prosodische Daten

Und an slavischen Korpora ...

- Gesprochene Korpora (im ČNK Einiges vorhanden; im Russ. aus den 1960ern)
- Spracherwerbskorpora (sehr geringen Mengen zu CZ und Russ., etwas mehr zu PL vorhanden)
- Diachrone Korpora (bisher nur einige Texte ohne Annotation in Titus) – Laufendes Arbeitsprojekt an der Uni Regensburg

Anhang: Weitere frei verfügbare slavische Korpora

- Textabfrage mit regulären Ausdrücken o.ä.
 - Oslo Corpus of Bosnian Texts
 - Russische Korpora des SFB 441, Tübingen (z. T.)
 - Slovakisches Nationalkorpus
 - Kroatisches Nationalkorpus
 - Serbisches Nationalkorpus
 - Polnisches Korpus des PWN-Verlags
- POS-annotierte Korpora
 - Tschechisches Nationalkorpus (disambiguiert)

- Polnisches IPIPAN-Korpus (ambig und disambiguiert)
- Russische Korpora des SFB 441, Tübingen (in Teilen; disambiguiert)
- Russisches Nationalkorpus (teilweise disambiguiert)
- Syntaktisch annotierte Korpora, Baumbanken
 - BulTreebank (HPSG)
 - Prague Dependency Treebank (Prager “Funktional-generative Deskription”)
 - (IPIPAN Testsuite für HPSG-Fragment des Polnischen)
- Links dazu unter

<http://www-slavistik.uni-regensburg.de/studium/links.html>
<http://www-slavistik.uni-regensburg.de/Corpus>
<http://www.sfb441.uni-tuebingen.de/c1/corpora.html>
<http://www.uni-tuebingen.de/uni/nss/docs/Korpora.html>

Literatur

- Błaszczak, J. (2001): *Investigation into the Interaction between the Indefinites and Negation*, Bd. 51 von *studia grammatica*. Akademie-Verlag, Berlin.
- Fanselow, G. (2004): *Against Remnant VP-Movement*. Universität Potsdam.
- Hajič, J./ P. Krbeč/ P. Květoň *et al.* (2001): *Serial Combination of Rules and Statistics: A Case Study in Czech Tagging*. In: *Association for Computational Linguistic, 39th Annual Meeting and 10th Conference of the European Chapter, Proceedings of the Conference*, 260–267. Morgan Kaufmann Publishers, Toulouse.
- Hajičová, E./ J. Panevová & P. Sgall (2000): *A Manual for Tectogrammatic Tagging of the Prague Dependency Treebank*. *Techn. Ber.*, UFAL/CKL, Charles University, Prague.
- Müller, G. (1998): *Incomplete Category Fronting*. Kluwer, Dordrecht.
- Przepiórkowski, A. (2004): *The IPIPAN Corpus. Preliminary Version*. Institute of Computer Science, PAS, Warszawa.