

1 はじめに

日本における漢文(古典中国語)の受容は、一つには訓読という形でおこなわれてきた。訓読は、言語処理という観点から見た場合、VO型の孤立語である古典中国語を、OV型の膠着語である日本語の読み下し文に変換する、という過程の一種だとみなせる。訓読を、返り点と送り仮名に分けたならば、VO型からOV型への変換を返り点が担い、孤立語から膠着語への変換を送り仮名が担っている、と考えることもできるだろう。

一方、コンピュータによる漢文の自動解析は、形態素解析→依存文法解析→直接構成素解析という手順によって、白文の統語構造を得ることができる(図1)、というのが、発表者の目論見^[1]である。入力された白文に対し、形態素解析によって、単語切りをおこなうと同時に、各単語の品詞を得る。依存文法解析によって、単語と単語の間の係り受け関係を解析すると同時に、文の切れ目を得る。直接構成素解析によって、各文の統語構造を解析木の形で得る。

ただし、コンピュータによる返り点の自動付与に関しては、直接構成素解析をおこなう必要は無く、依存文法解析までで(ほぼ)十分だ、というのが、本発表の主旨である。卑近な言い方をすれば、訓読の返り点は、漢文の統語構造ではなく、係り受け関係によって(ほぼ)記述可能だ、ということである。具体例として、本発表では、Universal Dependenciesによる漢文の依存文法解析と、富山房『漢文大系』における返り点を、特に「孟子定本」で比較し、依存文法解析と返り点を同一視できる文例、および、微妙に同一視できない文例について述べる。

2 Universal Dependencies による漢文の依存文法解析

発表者が班長を務める京都大学人文科学研究所共同研究班「東アジア古典文献コーパスの実証研究」(班員: ウィッテルン クリスティアン、守岡知彦、池田巧、山崎直樹、二階堂善弘、鈴木慎吾、師茂樹、李媛、白須裕之、藤田一乗)では、現在、漢文の依存文法解析に精力を傾注しており、その道具立ての一つとして、Universal Dependencies^[2](以下「UD」)の古典中国語への適用を研究している。依存文法解析それ自体は、Tesnièreの構造的統語

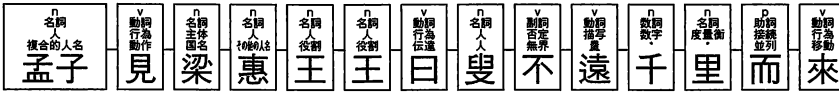
*京都大学人文科学研究所附属東アジア人文情報学研究センター

^[1]安岡孝一: 古典中国語(漢文)の依存文法解析と直接構成素解析, 漢字文献情報処理研究, 第18号(2018年10月), pp.56-62.

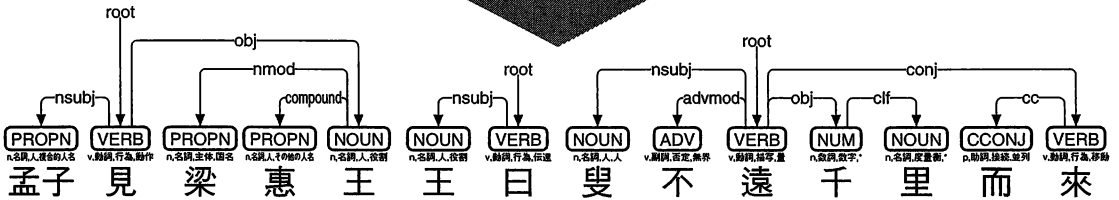
^[2]Joakim Nivre: Towards a Universal Grammar for Natural Language Processing, CICLing 2015: 16th International Conference on Intelligent Text Processing and Computational Linguistics (April 2015), pp.3-16.

孟子見梁惠王王曰叟不遠千里而來

形態素解析



依存文法解析



直接構成素解析

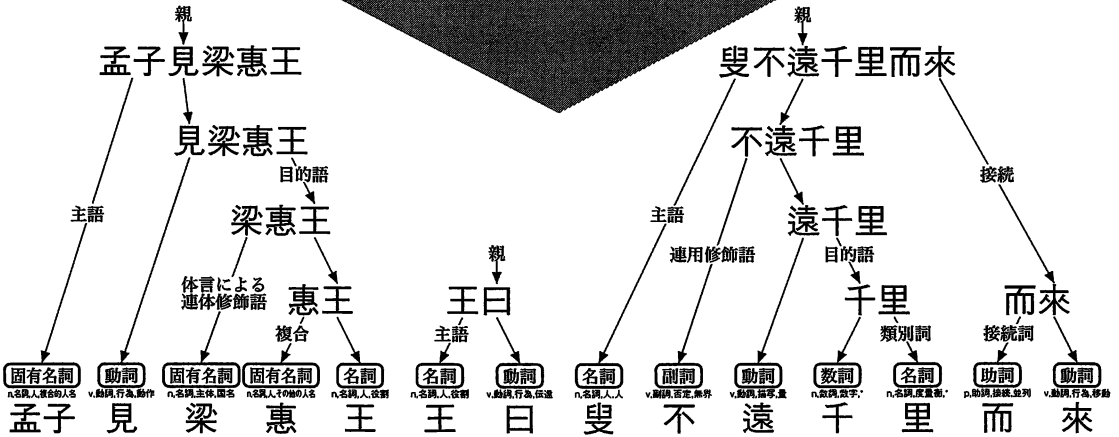


図 1: 白文「孟子見梁惠王王曰叟不遠千里而來」に対する統語構造解析^[1]の流れ

論^[3]に源を発し、Мельчукの有向グラフ記述^[4]によって、一応の完成を見た手法である。その最大の特長は、言語横断的な記述が可能だという点にあり、Мельчукの手法をコンピュータ向けに洗練したUDにおいても、言語に関わらない記述、という特長が前面に押し出されている。UDにおける文法構造記述は、句構造を考慮せず、全てを単語間のリンクとして表現する。これは、Мельчукの有向グラフ記述が、単語間のリンクという形態を取っていたからであり、そういう割り切りの結果として、言語横断的な文法構造記述を可能としているのである。

UDにおける係り受け関係の記述は、文中の単語をノードとする有向グラフにおいて、単語間の依存関係をリンクで表現する。各単語から出るリンクは複数ありうるが、各単語に入るリンクは必ず1本とする。また、リンクはループしない。リンクには、それぞれUD依存構造タグを付与する。古典中国語UDにおいては、表1に示す34種類^[5]のタグを使用している。タグのうち30種類は、もともとUDで規定されているものであり、4種類(nsubj:pass・csubj:pass・discourse:sp・flat:vv)は、その派生形である。rootはリンク元を持たないが、他のタグによるリンクは、リンク元の単語とリンク先の単語を1つずつ有する。たとえば、漢文の動賓構造は、動詞をリンク元、賓語をリンク先、とするobjというリンクで表現する。図2の「見」をリンク元とするobjリンクは、リンク先が3文字後の「王」となっており、「見」と「王」が動賓構造であることを示している。

なお、白文に対する依存文法解析は、その前段階として、単語切りという処理を必要と

表 1: 古典中国語に対する UD 依存構造タグ^[5]

	Nominals	Clauses	Modifier Words	Function Words
Core arguments	nsubj 主語 ↪nsubj:pass [受動文] obj 目的語 iobj 間接目的語	csubj 節主語 ↪csubj:pass [受動文] ccomp 節目的語 xcomp 節補語		
Non-core arguments	obl 斜格補語 vocative 呼称語 expl 形式語 dislocated 外置語	advcl 連用修飾節	advmod 連用修飾語 discourse 談話要素 ↪discourse:sp [文助詞]	aux 動詞補助成分 cop 繫辞 (copula) mark 標識 (marker)
Nominal dependents	nmod 体言による連体修飾語 nummod 数量による修飾語	acl 連体修飾節	amod 用言による連体修飾語	det 決定詞 clf 類別詞 case 格表示
Coordination	MWE	Loose	Special	Other
conj 接続 cc 接続詞	compound 複合 (endocentric) flat 並列 (exocentric) ↪flat:vv [動詞類]	list 細目 parataxis 隣接表現		root 親

^[3]Lucien Tesnière: *Éléments de Syntaxe Structurale*, Paris: C. Klincksieck (1959).
^[4]Igor A. Mel'čuk: *Dependency Syntax: Theory and Practice*, New York: State University of New York Press (1988).
^[5]安岡孝一: *Universal Dependencies にもとづく古典中国語 (漢文) の依存文法解析*, センター研究年報 2018 (2018 年 10 月).

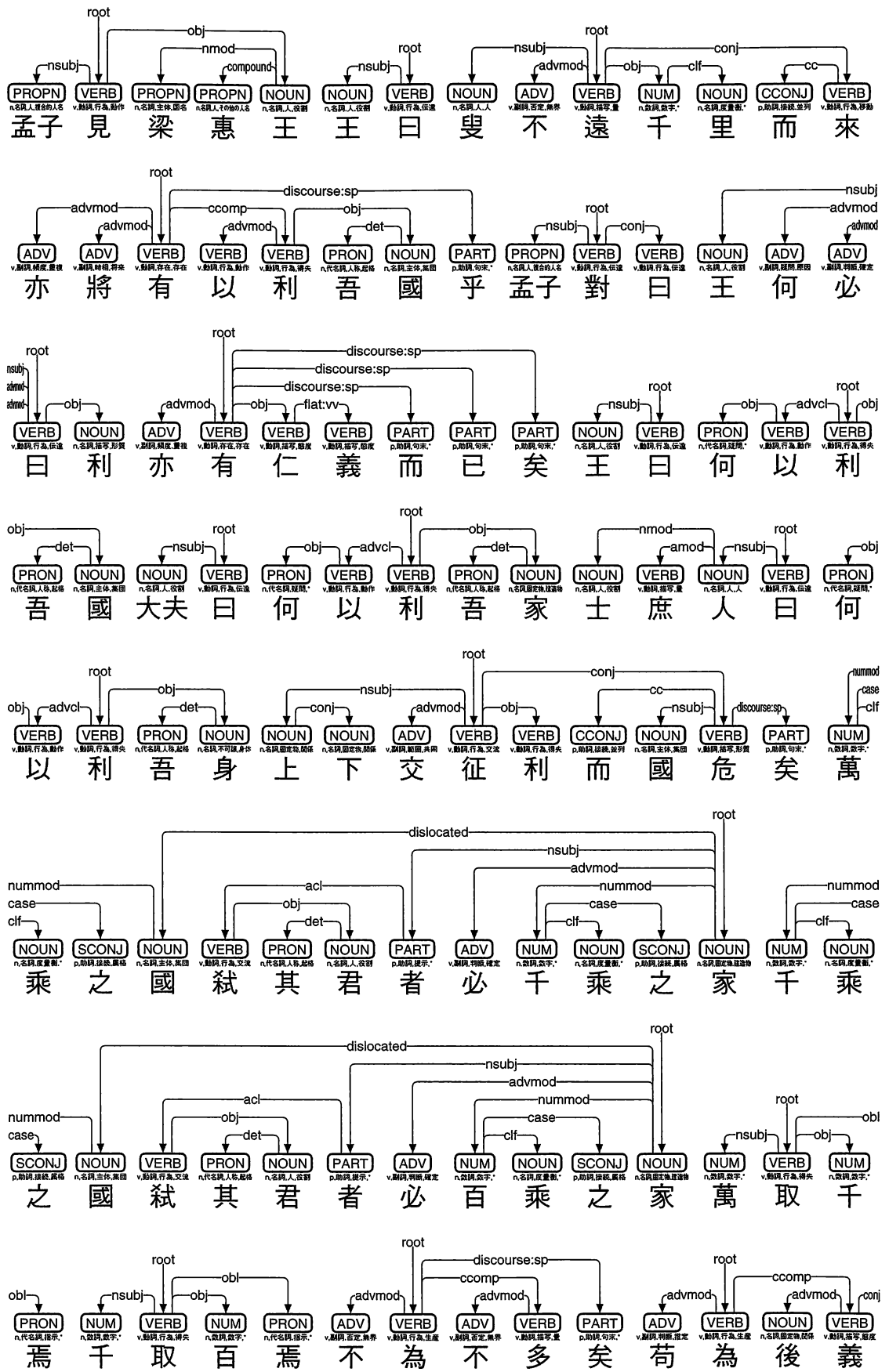


図 2: Universal Dependencies による依存文法解析結果 (「孟子定本」巻一冒頭)

する。白文では、単語と単語の間に区切りがないことから、単語というものを処理単位とする依存文法解析に際し、まず、白文を単語に区切る処理が必要となるのである。この処理を発表者は、漢文の形態素解析^[6]という形で実現し、白文の単語切りをおこなうと同時に、各単語に対して4階層の品詞を得ている(図1)。

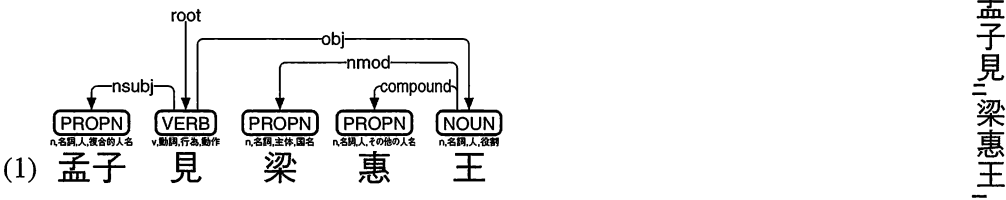
3 漢文の依存文法解析と返り点の関係

漢文の依存文法解析の結果から、いくつかのルールにもとづいて、訓読の返り点を導出したい。どのようなルールが必要となるのか、以下では「孟子定本」に対する依存文法解析結果(図2)をもとに、右向きのUDリンク、左向きのUDリンク、特別扱いすべき個別の文字、返り点の付け替え、の4つの視点から、返り点のルールを考えてみよう。なお、以下の文例では、返り元と返り先を視覚化すべく、レ点の代わりに一二点を用いている。各文例に対する富山房『漢文体系』での返り点は、図3を参照されたい。

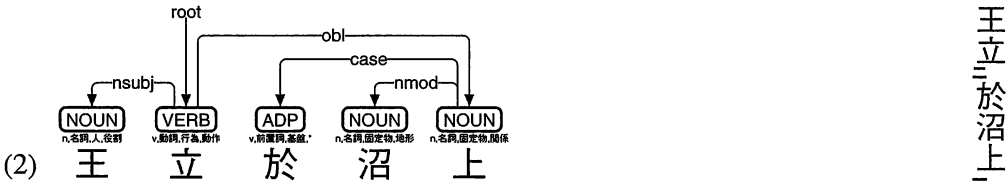
3.1 右向きのUDリンクに対する返り点のルール

右向きのUDリンクに対する返り点のルールは、以下のようなものが考えられる。

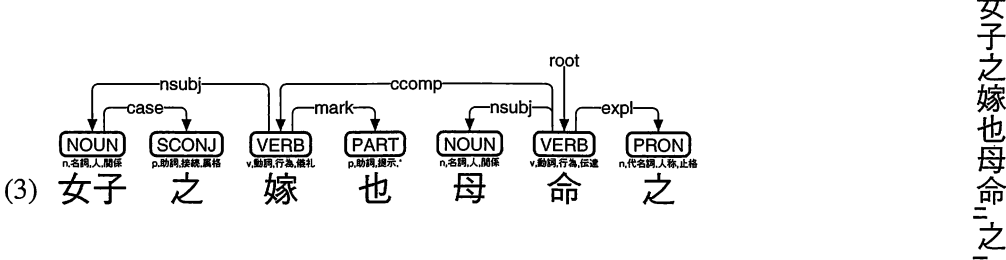
ルール① 右向きの obj リンクは、リンク先からリンク元へ返り点を打つ(文例1)。



ルール② 右向きの obl リンクは、リンク先からリンク元へ返り点を打つ(文例2)。



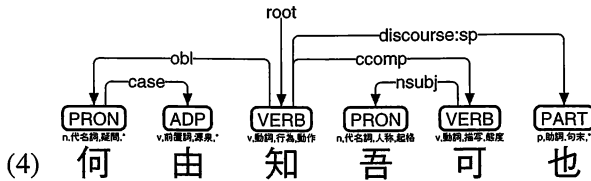
ルール③ 右向きの expl リンクは、リンク先からリンク元へ返り点を打つ(文例3)。



^[6]安岡孝一, ウィッテルン クリスティアン, 守岡知彦, 池田巧, 山崎直樹, 二階堂善弘, 鈴木慎吾, 師茂樹: 古典中国語(漢文)の形態素解析とその応用, 情報処理学会論文誌, Vol.59, No.2 (2018 年 2 月), pp.323-331.

ルール④ 右向きの ccomp および xcomp リンクは、リンク先からリンク元へ返り点を打つ (文例 4)。

何由知吾可也

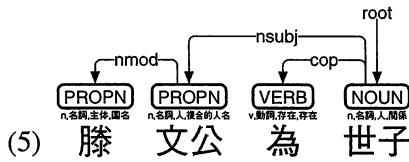


3.2 左向きの UD リンクに対する返り点のルール

左向きの UD リンクに対する返り点のルールは、以下のようなものが考えられる。

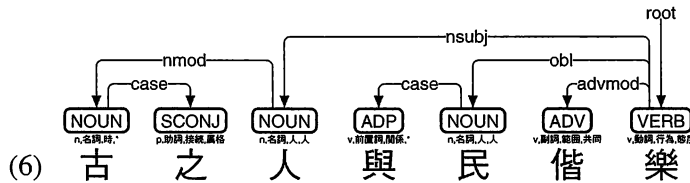
ルール⑤ 左向きの cop リンクは、リンク元からリンク先へ返り点を打つ (文例 5)。

滕文公為世子



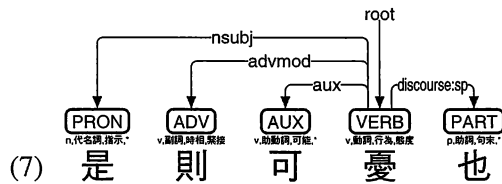
ルール⑥ 左向きの case および mark リンクは、リンク元からリンク先へ返り点を打つ (文例 6)。ただし、リンク先の形態素解析結果が「v, 前置詞, 基盤」の場合は、返り点を打たない (文例 2)。

古之人與民偕樂



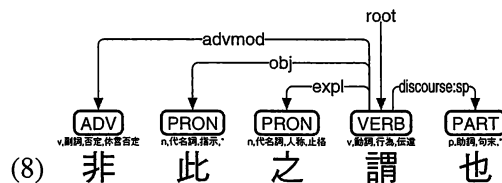
ルール⑦ 左向きの aux リンクは、リンク元からリンク先へ返り点を打つ (文例 7)。

是則可憂也

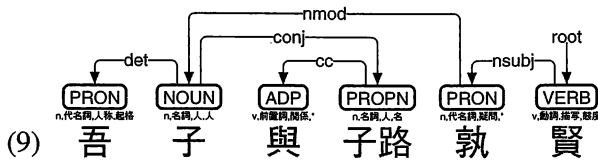


ルール⑧ 左向きの advmod リンクは、リンク先の形態素解析結果が「v, 副詞, 否定」「v, 副詞, 判断, 逆接」「v, 副詞, 時相, 将来」の場合、あるいはリンク先が「難」「易」の場合に限って、リンク元からリンク先へ返り点を打つ (文例 8)。

非此之謂也



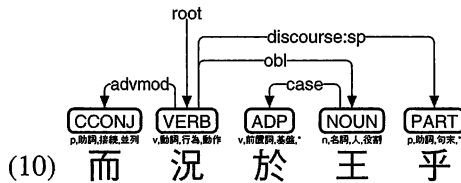
ルール⑨ 左向きの cc リンクは、リンク先の形態素解析結果が「v, 前置詞, 関係」の場合に限って、リンク元からリンク先へ返り点を打つ (文例 9)。



3.3 個別の文字に対するルール

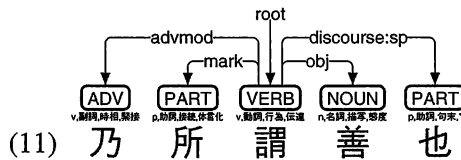
個別の文字に対する返り点のルールは、以下のようなものが考えられる。以下のルールの中には、富山房『漢文大系』の「孟子定本」における安井衡の「返りグセ」のようなものも含まれている気がするが、ルール化できそうなものは、できるだけルール化することにする。

ルール⑩ ルール①～④で打った返り点において、返り先が「況」の場合、返り点を削除する。ただし、返り点の返り元から、左向きの case あるいは mark リンクが出ている場合は、返り点を削除する代わりに、返り先をそのリンク先に移動する。



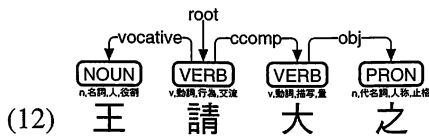
文例 10 では、ルール②にもとづいて「王」から「況」へ返り点を打つべきところ、返り先を「於」へ移動している。このルール⑩は、動詞である「況」を、「いわんや」と訓読するためのものである。

ルール⑪ ルール①～④で打った返り点において、返り先が「謂」であり、かつ、その「謂」から「所」へ左向きの mark リンクが出ている場合、「謂」を返り先とする返り点を削除する。



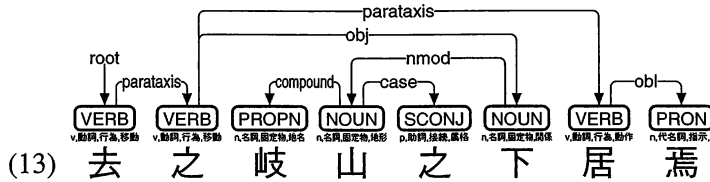
文例 11 では、「善」から「謂」への返り点 (ルール①) を削除している。このルール⑪は、動詞である「謂」に対し、「所謂」を「いわゆる」と訓読するためのものである。

ルール⑫ ルール①あるいは④で打った返り点において、返り先が「請」であり、かつ、その「請」から vocative リンクが出ている場合、返り点を削除する。



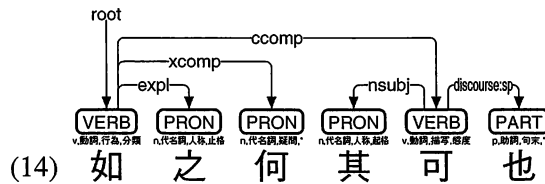
文例 12 では、「大」から「請」への返り点(ルール④)を削除している。このルール⑫は、呼びかけを意味する「請」において、訓読を命令調にするためのものである。

ルール⑬ ルール②で打った返り点において、返り元が「焉」であり、かつ、返り先の形態素解析結果が「v, 動詞, 描写」以外の場合、返り点を削除する。



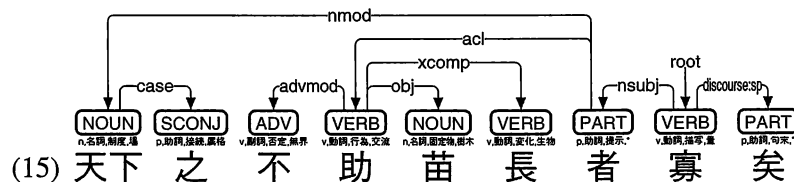
文例 13 では、「焉」から「居」への返り点を削除している。このルール⑬は、「於之」を意味する代名詞「焉」に対し、比較を意味する場合を除いて、訓読しないためのものである。

ルール⑭ ルール④で打った返り点において、返り先が「如」であり、かつ、その「如」から obj リンクもしくは expl リンクが出ている場合、返り点を削除する。



文例 14 では、「何」から「如」への返り点と、「可」から「如」への返り点を、いずれも削除している。

ルール⑮ ルール④で打った返り点において、返り先が「助」であり、かつ、xcomp リンクによる場合、返り点を削除する。加えて、その「助」が他のルールによる返り点の返り元である場合、返り元を、削除した返り点の返り元へ移動する。



文例 15 では、ルール①にもとづいて「苗」から「助」へ、ルール④にもとづいて「長」から「助」へ、ルール⑧にもとづいて「助」から「不」へ、それぞれ返り点を打つべきところ、「長」から「助」への返り点は削除している。加えて、「助」から「不」への返り点は、返り元を「長」へ移動している。

勸_二齊_一伐_二燕_一有_二諸_一



所以別野人也

所以別野人也



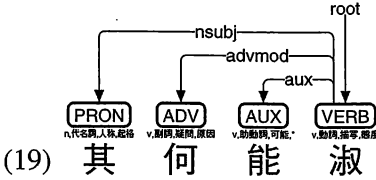
所以別野人也

所以自為一則吾不知二也



所以自為一則吾不知二也

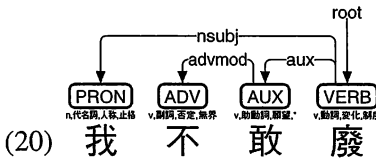
ルール⑱ ルール⑦で打った返り点において、返り先が「能」であり、かつ、その「能」が他のルールによる返り点の返り元でない場合、「能」を返り先とする返り点を削除する。



其何能淑

文例 19 では、「淑」から「能」への返り点を削除している。このルール⑱は、助動詞である「能」を、「よく～す」と訓読するためのものである。ただし、「不能」を「あたわず」と訓読する場合は、返り点を削除しない。

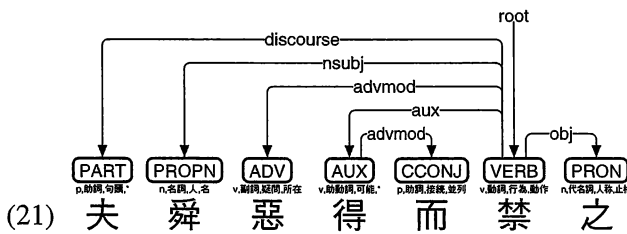
ルール㉑ ルール⑦で打った返り点において、返り先が「敢」の場合、返り点を削除する。加えて、その「敢」が他のルールによる返り点の返り元である場合、返り元を、削除した返り点の返り元へ移動する。



我不敢廢

文例 20 では、ルール⑦にもとづいて「廢」から「敢」へ、ルール⑧にもとづいて「敢」から「不」へ、それぞれ返り点を打つべきところ、削除と移動で「廢」から「不」への返り点としている。このルール㉑は、助動詞である「敢」を、「あえて～す」と訓読するためのものである。

ルール㉒ ルール⑦で打った返り点において、返り先が「得」であり、かつ、その「得」から「而」へ右向きの advmod リンクが出ている場合、返り点を削除する。加えて、その「得」が他のルールによる返り点の返り元である場合、返り元を、削除した返り点の返り元へ移動する。



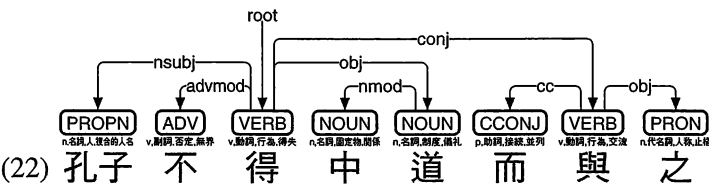
夫舜惡得而禁之

文例 21 では、「禁」から「得」への返り点を削除している。このルール㉒は、助動詞である「得而」を、動詞として訓読するためのものである。

3.4 返り点の付け替えに対するルール

ここまでのルールで打った返り点に対し、以下では、返り点を削除・移動・追加するルールを考える。

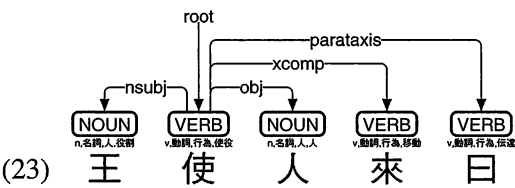
ルール②② ルール①～②①で打った返り点の返り元から、右向きの conj・clf・flat・case リンクが出ている場合、それらのリンク先のうち文末に最も近いものへ、返り元を移動する。



孔子不_上得_上中道而與_上之_上

文例 22 では、ルール①にもとづいて「道」から「得」へ、ルール①にもとづいて「之」から「與」へ、ルール⑧にもとづいて「得」から「不」へ、それぞれ返り点を打つべきところ、「得」から「不」への返り元を「與」へ移動している。

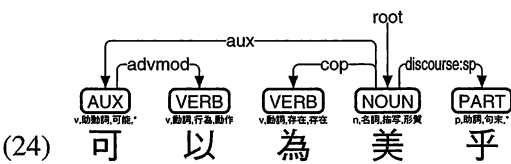
ルール②③ ルール①～②②で打った返り点において、1 つの返り先へ複数の返り元から返り点が集中している場合、それらの返り元のうち文末に最も近いものを残し、他の返り元は削除する。



王使_上人來_上曰_上

文例 23 では、ルール①にもとづいて「人」から「使」へ、ルール④にもとづいて「來」から「使」へ、それぞれ返り点を打つべきところ、「人」からの返り点は削除している。

ルール②④ ルール①～②③で打った返り点において、1 つの返り元から複数の返り先へ返り点がある場合、それらの返り先のうち文頭に最も近いもの以外に返り元を追加し、返り点を後ろから順に辿る形とする。



可_上以_上為_上美_上乎_上

文例 24 では、ルール⑤にもとづいて「美」から「為」へ、ルール⑦にもとづいて「美」から「可」へ、それぞれ返り点を打つべきところ、「為」に返り元を追加して、「美」「為」「可」を順に辿る返り点としている。

- (1) 孟子見_レ梁惠王_一〔卷一〕
- (2) 王立_二於沼上_一〔卷二〕
- (3) 女子之嫁也母命_レ之〔卷六二〕
- (4) 何由知_二吾可_一也〔卷一七〕
- (5) 滕文公爲_二世子_一〔卷五一〕
- (6) 古之人與_レ民偕樂〔卷一二〕
- (7) 是則可_レ憂也〔卷八二八〕
- (8) 非_二此之謂_一也〔卷四二〕
- (9) 吾子與_二子路_一孰賢〔卷三一〕
- (10) 而況於_二王乎_一〔卷四九〕
- (11) 乃所_レ謂善也〔卷十一六〕
- (12) 王請大_レ之〔卷二三〕
- (13) 去之_二岐山之下_一居焉〔卷二一四〕
- (14) 如_レ之何其可也〔卷二一一〕
- (15) 天下之不_二助_一苗長_二者寡矣_一〔卷三二〕
- (16) 勸_レ齊伐_レ燕有_レ諸〔卷四八〕
- (17) 所_二以別_一野人也〔卷五三〕
- (18) 所_二以自爲_一則吾不_レ知也〔卷四五〕
- (19) 其何能淑〔卷七九〕
- (20) 我不_二敢廢_一〔卷八二四〕
- (21) 夫舜惡得而禁_レ之〔卷十三三三〕
- (22) 孔子不_下得_二中道_一而與_レ之〔卷十四三三〕
- (23) 王使_二人來_一曰〔卷四二〕
- (24) 可_二以爲_一美乎〔卷十一八〕
- (25) 叟不_レ遠_二千里_一而來〔卷一〕
- (26) 未_二嘗與_一之言_二行事_一也〔卷四六〕
- (27) 請以_レ戰喻〔卷一三〕
- (28) 以_レ齊王由_レ反_レ手_レ也〔卷三一〕
- (29) 蓋_二亦反_一其本_二矣_一〔卷一七〕
- (30) 曰_下其取_二諸民_一之不義也_上〔卷十四〕

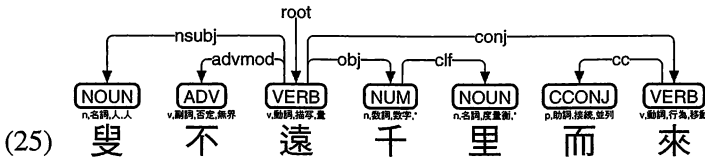
図 3: 富山房『漢文大系』の「孟子定本」における返り点の文例

3.5 例外

ここまでのルールを適用することで、「孟子定本」の依存文法解析結果のうち、9割以上が富山房『漢文大系』の返り点と一致する。ただ、数パーセント例外があり、それらに関しては富山房『漢文大系』と微妙に一致しない。以下では、そのようないくつかの例外について述べる。

3.5.1 「叟不遠千里而來」

文例 25 は、ルール①・⑧・②を適用することで、右図のような返り点を得られる。



叟
不
遠
千
里
而
來

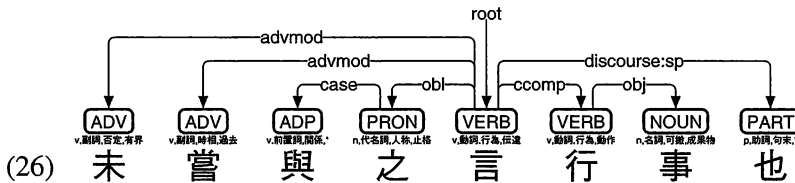
しかし『漢文大系』(図 3(25))では、「不」への返り点が、「來→不」ではなく「遠→不」となっている。端的には、ルール⑧にもとづく「遠」から「不」への返り点に対し、右向きの conj によるルール②が適用されていない。これは何を意味するのか。

一般に「不 X 而 Y」という文型においては、「不」が X だけを否定しているのか、あるいは X と Y の両方を否定しているのか、文型として曖昧である。ルール②は、このような文型に対し、X と Y の両方を否定しているとみなして、返り元の移動をおこなっており、少なくとも文例 22 に関しては、正しい返り点を導出している。その一方、文例 25 に関しては、実は X だけを否定する文例であることから、ルール②を適用すると、返り点を間違ってしまうわけである。

しかしながら「不 X 而 Y」という文型において、「不」が X だけを否定しているのか、あるいは X と Y の両方を否定しているのかは、文の内容や前後関係を読まなければ、本質的に決めることができない。その意味で「不 X 而 Y」という文型における返り点は、依存文法解析という「表層的」な解析手法の限界を超えており、本発表の射程においては、例外とせざるを得ないということである。

3.5.2 「未嘗與之言行事也」

文例 26 は、ルール①・④・⑥・⑧を適用することで、右図のような返り点を得られる。

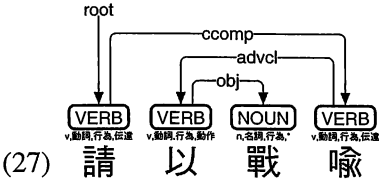


未
嘗
與
之
言
行
事
也

しかし『漢文大系』(図 3(26))では、「事」からの返り点が、「事→行→言→未」ではなく「事→言→未」となっている。端的に言えば、「事を行う」ではなく、「行事」という熟語として訓読している。しかしながら、どの 2 文字(あるいは 3 文字以上)を熟語として読むかは、本質的に決定不可能であり、例外として扱わざるを得ないだろう。

3.5.3 「請以戰喻」

文例 27 は、ルール①・④を適用することで、右図のような返り点を得られる。

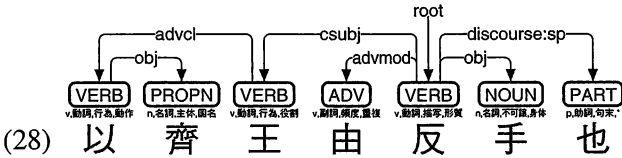


請
下
以
戰
上
喻

しかし『漢文大系』(図 3(27))では、「喻→請」の返り点がない。これは端的に言って、ルール⑫の不備である。ルール⑫では、呼びかけを意味する「請」を返り先とする返り点を削除している。しかしながら、呼びかけであることを判定するのに **vocative** リンクを用いているため、いきなり「請」から始まる文例 27 では判定に失敗する。このような文において、呼びかけであることを判定するためには、文の内容や前後関係を読まなければ無理だが、それは依存文法解析の限界を超えていると考えられる。

3.5.4 「以齊王由反手也」

文例 28 は、ルール①を適用することで、右図のような返り点を得られる。

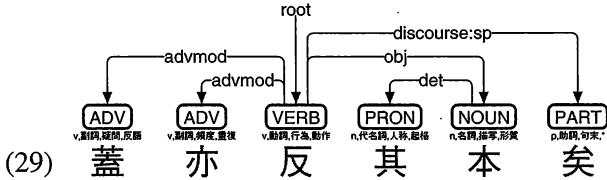


以
齊
王
由
反
手
也

しかし『漢文大系』(図 3(28))では、「手→反」の返り点が「手→反→由」となっており、「反」から「由」への返り点が増えている。これは「由」を再読文字とみなして「なお～のごとし」と訓読しているためであり、「孟子定本」における「由」の用法としても、かなり例外的な返り点である。

3.5.5 「蓋亦反其本矣」

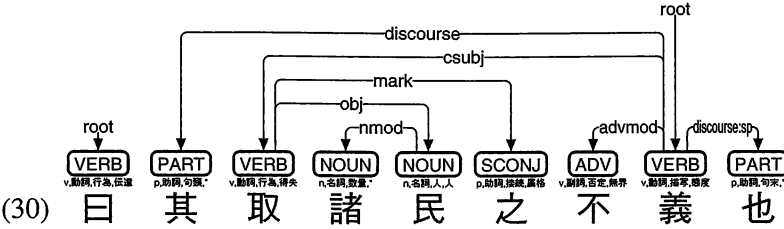
文例 29 は、ルール①を適用することで、右図のような返り点を得られる。



蓋
亦
反
其
本
矣

しかし『漢文大系』(図 3(29))では、「本→反」の返り点が「本→反→蓋」となっており、「反」から「蓋」への返り点が増えている。「蓋」を再読文字とみなして「なんぞ～ざる」と訓読しているためだが、これを「けだし」と分離してルール化するのは難しく、現時点では例外とせざるを得ない。

文例 30 は、ルール①・⑧を適用することで、右図のような返り点が得られる。



しかし『漢文大系』(図 3(30))では、「也」から「曰」への返り点が増えており、「義」から「不」への返り点が無い。「義→不」については、これを「不義」という熟語として訓読しているためである。一方、「曰」への返り点は、非常に難しい問題を孕んでいる。『漢文大系』の「孟子定本」においては、地の文に現れる「曰」については返り点が無いが、会話文の中に現れる「曰」には返り点がしばしば打たれていて、会話文の中の会話文に対し「～と曰う」と訓読している。しかしながら、ある「曰」が、地の文に現れているのか、会話文の中に現れているのかを判定するには、文どうしの入れ子構造を解析しなければならず、それは依存文法解析の限界を遥かに超えていると考えられる。

4 おわりに

本発表では、「孟子定本」を Universal Dependencies によって依存文法解析し、解析結果を富山房『漢文大系』の返り点と比較した。ここから 24 のルールを導出し、漢文の依存文法解析結果を、返り点へと変換する手法の可能性について、議論をおこなった。これらのルールにもとづけば、漢文の依存文法解析結果の 9 割以上を、返り点と対応させることができる。

一方、前節で議論した例外は、本手法の限界、ひいては漢文の依存文法解析の限界を浮かび上がらせる結果となった。これらの例外をも自動処理するためには、漢文の依存文法解析だけでは不十分であり、訓読結果としての日本語(読み下し文)に対する依存文法解析との突合、それにもとづく適切な送り仮名の自動付与、さらには漢文の意味解析が必要となるだろう。これらは端的には、テキストの解釈そのものに入り込んでいくことを意味しており、かなり困難な作業が予想される。返り点には、それだけ多く先人の智慧が埋め込まれている、ということなのであろう。

なお、本発表で示した 24 のルール(例外は除く)を、形態素解析エンジン MeCab^[7]および依存文法解析エンジン UDPipe^[8]とともに、以下の URL で試しに実装した。「孟子定

^[7]Taku Kudo, Kaoru Yamamoto, Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (July 2004), pp.230-237.

^[8]Milan Straka and Jana Straková: Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe, Proceedings of the CoNLL 2017 Shared Task (August 2017), pp.88-99.

本」に特化して解析エンジンを鍛えてしまったため、「孟子定本」以外の漢文は今一つ読みこなすことができないのだが、よければ「解析」ボタンを2回以上、押してみてください。

<http://kanji.zinbun.kyoto-u.ac.jp/~yasuoka/kyodokenkyu/2018-12-01.html>

「漢文の依存文法解析と返り点の関係について」試作サイト

未_レ有_レ義_レ而_レ後_レ其_レ君_レ者_レ也

解析

未_レ有_レ義_レ而_レ後_レ其_レ君_レ者_レ也

未_レ有_レ義_レ而_レ後_レ其_レ君_レ者_レ也

未_レ有_レ義_レ而_レ後_レ其_レ君_レ者_レ也

未_レ有_レ義_レ而_レ後_レ其_レ君_レ者_レ也

未_レ有_レ義_レ而_レ後_レ其_レ君_レ者_レ也

未_レ有_レ義_レ而_レ後_レ其_レ君_レ者_レ也