

古典中国語 Universal Dependencies への挑戦

安岡 孝一^{1,a)} ウィッテルン クリスティアン^{1,b)} 守岡 知彦^{1,c)} 池田 巧^{1,d)} 山崎 直樹^{2,e)}
二階堂 善弘^{2,f)} 鈴木 慎吾^{3,g)} 師 茂樹^{4,h)}

概要：古典中国語(漢文)の解析手法として、Universal Dependencies の古典中国語への適用に挑戦した。言語横断的な依存構造記述である Universal Dependencies は、ニューラルネットを用いた言語解析ツール(特に SyntaxNet)等に採用されていることから、係り受け解析への応用が容易だと考えられる。しかしながら、古典中国語の文法構造は、Universal Dependencies とは必ずしも相性が良くなく、いくつかの点で齟齬が生じている。この点を含め、現時点での古典中国語 Universal Dependencies の検討状況を、他言語 Universal Dependencies との比較も含め、大まかに報告する。

キーワード：漢文コーパス、係り受け解析、依存構造解析

1. はじめに

古典中国語(漢文)テキストをコンピュータ処理するためには、白文(単なる漢字の列)のままではどうにもならず、テキストを自然言語解析する必要がある。たとえば現代の欧米諸語であれば、テキストは単語単位に区切られており、文末を表す記号も付加されていることが多いので、語彙抽出などをおこなうのは比較的容易である。一方、古典漢文においては、単語の間にも文の間にも区切りを持たない白文のままでは、せいぜい文字列検索しかおこなえない。古典漢文のような書写言語の解析においては、まず、単語を認識した上で単語間の関係について解析をおこなう必要がある、そのためには形態素解析をおこなった上で、係り受け解析へとコマを進めていかねばならない。

この問題に対し、われわれは、2008年4月より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研究」を組織し、さらに2013年4月より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応用研究」を組織して、古典漢文に対する形態素解析の研究をお

こなってきた[1]。その上で、2016年4月より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」を組織して、古典漢文に対する係り受け解析の研究に着手した。その過程において、われわれは、Universal Dependencies [2] (以下、UD) という、言語横断的な依存構造記述に関わらざるを得なくなった。というのも、時おりしも google から SyntaxNet が発表 [3] され、言語横断的な文法解析の可能性が広がったが、この SyntaxNet が UD に基づくものだったからである。

本稿では、UD の古典漢文への適用に関して、われわれの試行錯誤の現状を、ざっと述べることにする。

2. UD2.0 の概要と CoNLL-U フォーマット

UD は、言語横断的な依存構造記述を目指すプロジェクトである。端的には、品詞・形態素属性・依存構造情報を、言語に依存せず記述することで、言語横断的な情報とすることを目指している。句構造を考慮せずに係り受け関係を

表1 CoNLL-U フォーマットのフィールド

ID	文中での単語番号
FORM	単語
LEMMA	単語の正規形
UPOSTAG	UD 品詞
XPOSTAG	言語依存の品詞
FEATS	UD 形態素属性
HEAD	リンク元の単語番号
DEPREL	UD 依存構造
DEPS	拡張依存関係
MISC	その他

¹ 京都大学 Kyoto University
² 関西大学 Kansai University
³ 大阪大学 Osaka University
⁴ 花園大学 Hanazono University
^{a)} yasuoaka@kanji.zinbun.kyoto-u.ac.jp
^{b)} wittern@zinbun.kyoto-u.ac.jp
^{c)} tomo@kanji.zinbun.kyoto-u.ac.jp
^{d)} ikeda@zinbun.kyoto-u.ac.jp
^{e)} ymzknk@kansai-u.ac.jp
^{f)} nikaido@kansai-u.ac.jp
^{g)} suzukish@lang.osaka-u.ac.jp
^{h)} s-moro@hanazono.ac.jp

1	孟子	孟子	PROPN	n, 名詞, 人, 複合的人名	-	2	nsubj	-	SpaceAfter=No
2	見	見	VERB	v, 動詞, 行為, 動作	-	0	root	-	SpaceAfter=No
3	梁	梁	PROPN	n, 名詞, 主体, 国名	-	5	nmod	-	SpaceAfter=No
4	惠	惠	PROPN	n, 名詞, 人, その他の人名	-	5	compound	-	SpaceAfter=No
5	王	王	NOUN	n, 名詞, 人, 役割	-	2	obj	-	SpaceAfter=No

図1 「孟子見梁惠王」の CoNLL-U データ

記述すべく、全ての構文構造を「単語」間の依存構造で記述するのが特徴である(付録参照)。

現時点の UD2.0 は、各単語ごとに 10 個のフィールド(表 1) からなる CoNLL-U フォーマットを定義しており、この CoNLL-U フォーマットを用いて、文を記述する。文字コードは UTF-8 であり、各フィールドの区切りは「タブ」U+0009 が 1 つ、各行(各単語)の区切りは「改行」U+000A が 1 つである。空行(2 つ以上の連続する「改行」)は、文の区切りを意味する。

ID は 1 から始まる自然数である。ただし、複数の単語にまたがる記述をおこないたい場合には、1-3 のような「範囲」を ID に書くこともできる。また、5.1 のような「小数」により、5 と 6 の間に「隠れている単語」を記述することも可能となっている。FORM と LEMMA は、空白を含みうるので、注意が必要である。UPOSTAG は 17 の品詞が、FEATS は 21 の形態素属性が、DEPREL は 37 の依存構造タグが、現時点では定義されている。HEAD と DEPREL は、内容語の依存構造に特化して設計されており、各単語のリンク元は高々 1 つしか記述できない。ただし、どうしても複数の単語からの依存関係を記述したい場合には、一応、DEPS による拡張も許されている。

3. UD2.0 の古典中国語への適用

われわれはこれまで、古典漢文に対し、MeCab を用いた形態素解析の研究をおこなってきた [1]。その成果をもとに、係り受け解析への足掛かりとすべく、UD2.0 の古典漢文への適用に挑戦してみた。具体的には、Pulleyblank の古典中国語文法書 [4] の各例文について、どのように UD2.0 を適用していくべきか検討・考察をおこないつつ、実際に CoNLL-U データの構築*1 をおこなっている。なお、検討・考察においては、現代中国語への UD の適用 [5] も参考にしたが、方針が必ずしも一致しているわけではない。

3.1 ID・FORM・LEMMA

われわれは、MeCab 漢文形態素解析のこれまでの成果を、漢文の単語分割にそのまま用いることにした。たとえば「孟子見梁惠王」という文は、形態素解析により「孟子」「見」「梁」「惠」「王」という 5 つの単語に分割し、それぞれ 1~5 の ID を振った上で、各単語を FORM に入れる(図

1)。ID に「範囲」や「小数」は用いない。LEMMA は、対応する FORM の「康熙字典体」ということにして、ある程度「正規形」っぽくしたが、必ずしも徹底できているわけではない。

3.2 UPOSTAG

われわれの MeCab 漢文形態素解析においては、大品詞・品詞・意味素性・小素性の 4 階層からなる品詞階層を用いている。この品詞階層のうち、大品詞と品詞の組み合わせが、ほぼ UD 品詞に対応しているように感じられることから、UPOSTAG の 17 の品詞のうち、NOUN・PRON・NUM・VERB・ADP・ADV・AUX・PART・INTJ の 9 つについて、表 2 に示す形で対応させてみることにした。ただし、表中の特例に示す通り、「n, 名詞」のうち「n, 名詞, 人, 姓氏」など 6 種類については、固有名詞である可能性が高いことから、PROPN に対応させておくことにした。また、「p, 助詞, 接続, 属格」と「p, 助詞, 接続, 並列」については、それぞれ SCONJ と CCONJ に対応させた。これらに加え、PUNCT と SYM は、われわれの MeCab 漢文形態素解析においても、特殊用途として「s, 記号」と「s, 文字」を準備していることから、一応、これらに対応しておいた。

残る UD 品詞 3 つ (ADJ・DET・X) について、われわれの検討結果を以下に述べる。まず ADJ については、古典中国語において、われわれは形容詞という分類を廃止 [6] し

表 2 MeCab 漢文形態素解析の品詞階層と UD 品詞

大品詞, 品詞	UPOSTAG	特例	
n, 名詞	NOUN	—, 人, 姓氏	} PROPEN
		—, 人, 名	
		—, 人, その他の人名	
		—, 人, 複合的人名	
		—, 主体, 国名	
		—, 固定物, 地名	
n, 代名詞	PRON		
n, 数詞	NUM		
v, 動詞	VERB		
v, 前置詞	ADP		
v, 副詞	ADV		
v, 助動詞	AUX		
p, 助詞	PART	—, 接続, 属格	SCONJ
		—, 接続, 並列	CCONJ
p, 感嘆詞	INTJ		
s, 記号	PUNCT		
s, 文字	SYM		

*1 CoNLL-U エディタの α 版を、<http://kanji.zinbun.kyoto-u.ac.jp/~yasuoka/kyodokenkyu/2017-12-15/conllusvg/meca-b-kanbun.html> に公開しているので参考にされたい。

表3 古典中国語に対する UD 依存構造タグ

	Nominals	Clauses	Modifier Words	Function Words
Core arguments	nsubj 主語 ↔nsubj:pass 主語 [受動文] obj 目的語 iobj 間接目的語	csubj 節主語 ↔csubj:pass 節主語 [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 談話要素 [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	flat 名詞並列 compound 名詞複合	parataxis 隣接表現		root 親

ており、したがって ADJ も使用しない。同様に DET は、古典中国語においては PRON の一部の用法に過ぎず、したがって DET も使用すべきでは無いと考えられる。X (それ以外の品詞) は、古典中国語には無いと考えられる。

3.3 XPOSTAG・FEATS

UD2.0 においては、FEATS には 21 の形態素属性が準備されている。しかし、この形態素属性は、われわれの意味素性・小素性とは、かなり異なっている。たとえば、「n, 名詞, 固定物, 地名」には FEATS の NameType=Geo が対応しそうだが、「n, 名詞, 主体, 国名」に対応する FEATS の属性は、NameType=Geo と NameType=Nat にまたがりそうであり、そのどちらになるかは文脈依存である。FEATS の形態素属性に、われわれの意味素性・小素性を自動変換するのは、非常な困難を伴う。そこで現時点では、FEATS は使わないこととし、われわれの小品詞・品詞・意味素性・小素性を、そのまま XPOSTAG に入れている。

3.4 HEAD・DEPREL・DEPS

DEPREL には、UD 依存構造タグのうち、古典中国語に特化した 29 種類と、その派生形 3 種類 (nsubj:pass と csubj:pass と discourse:sp) を使用することにした。これら 32 種類の UD 依存構造タグ (表 3) は、Pulleyblank の古典中国語文法書 [4] の各例文について、手作業で UD 依存構造を決定していくにあたり、記述に必要となったタグである。なお、各単語の依存先 (リンク元) は、HEAD を用いて 1 つに限定し、DEPS は使用しないこととした。

たとえば、図 1 の「孟子見梁惠王」においては、「見」を root とし、そこから「孟子」を nsubj、「王」を obj とする依存構造リンクを記述している。さらに「王」からは、「惠」を compound、「梁」を nmod とする依存構造リンクを記述しており、全体として依存構造ツリー (図 2) を構成している。

UD においては、コンピュータは内容語とせず、コンピュータの補語を root とする。この点は、古典中国語 UD においても同様とした。たとえば、「孟子為卿於齊」においては、「卿」を root とし、そこから「孟子」を nsubj、「為」を cop とする依存構造リンクを記述した (図 3)。ただし、「也」を句末におくコンピュータ文については、現代中国語 UD に倣って、「也」を cop ではなく discourse:sp で繋いだ。たとえば、「此文王之勇也」においては、「勇」を root とし、そこから「此」を nsubj、「也」を discourse:sp とする依存構造リンクを記述した (図 4)。

古典中国語においては、主語を伴わない文も数多く見られる。そのような文も、UD は自然に記述可能である。たとえば、「則不能安子思」においては、「安」を root とし、そこから「子思」を obj、「能」を aux とする依存構造リンクを記述した (図 5)。なお、「不」は「能」の advmod とし、「則」は「安」の advmod としている。

古典中国語における兼語文 (動詞の目的語が次の動詞の主語を兼ねる文で、使役動詞などに見られる) に対しては、動詞どうしの依存関係を xcomp で記述することにした。たとえば、「王使人來曰」においては、「人」を「使」の obj とした上で、「來」を「使」の xcomp とする依存構造リンクを記述した (図 6)。なお、「曰」については、この文だけだとそもそも誰が言ったのかすら判然とせず、ざっくり「使」の parataxis としている。

「若〜何」形式の疑問文に対しては、「若」と「何」を xcomp で繋ぐことにした。たとえば、「子若國何」においては、「國」を「若」の obj とした上で、「何」を「若」の xcomp とする依存構造リンクを記述した (図 7)。あるいは、「若之何子之不言也」においては、「何」を「若」の xcomp、形式目的語の「之」を expl とした上で、実際の目的語「子之不言」の「言」を ccomp とする依存構造リンクを記述した (図 8)。

一方、「p, 助詞, 提示」にあたる単語については、UD 依

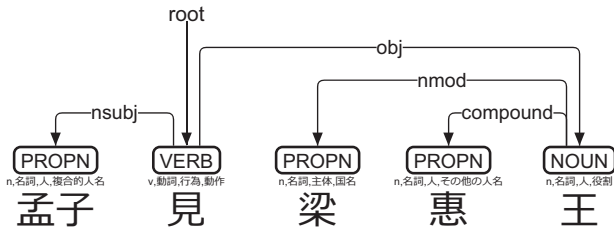


図2 「孟子見梁惠王」のUD依存構造

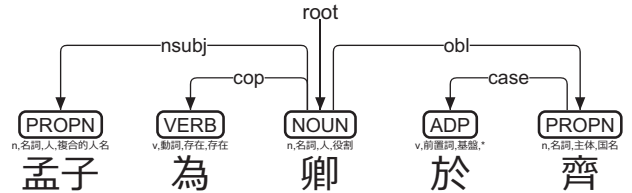


図3 「孟子為卿於齊」のUD依存構造

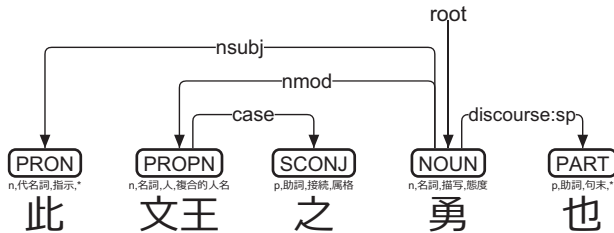


図4 「此文王之勇也」のUD依存構造

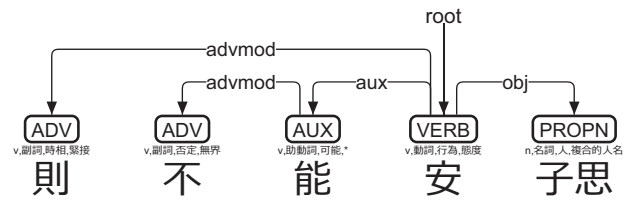


図5 「則不能安子思」のUD依存構造

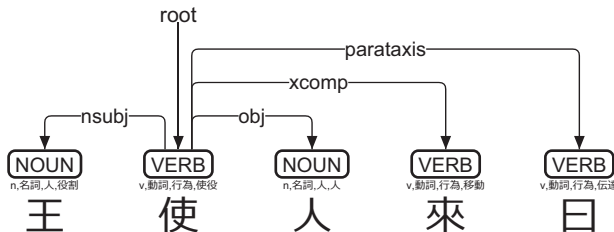


図6 「王使人來曰」のUD依存構造

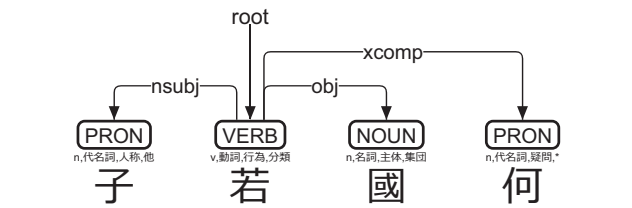


図7 「子若國何」のUD依存構造

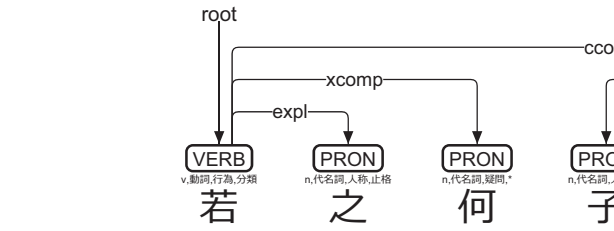


図8 「若之何子之不言也」のUD依存構造

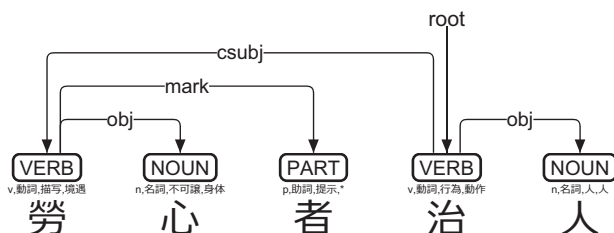
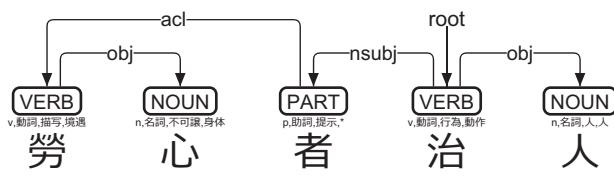


図9 「勞心者治人」のUD依存構造(二案)

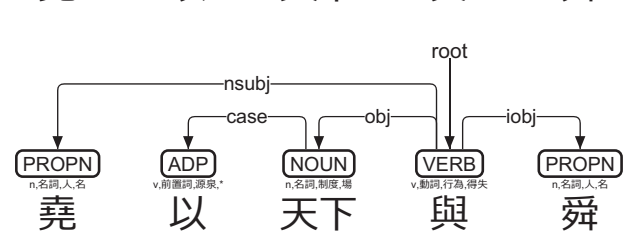
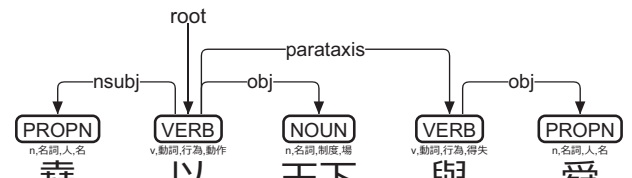


図10 「堯以天下與舜」のUD依存構造(二案)

存構造をどう決定すべきか、悩ましい場面も少なくない。例として「勞心者治人」の「者」を見てみよう。この文では、「勞心」という動賓構造を「者」で体言化して主語としているわけだが、この「者」を *nsubj* として扱うか、あるいは *mark* として扱うかによって、UD 依存構造が全く変わってしまう(図9)。これが仮に「勞心者」ではなく、たとえば「死者」であった場合を考えると、*nsubj* の方が良いようにも思える*2ののだが、それが UD の考え方に合致しているのかどうか、かなり悩ましいところである(付録参照)。

また、「v, 前置詞」については、そもそも「v, 動詞」との間で議論があり、UD 依存構造を決定する際に、その問題があらわになってきている。例として「堯以天下與舜」の「以」を見てみよう。「以」は元々は動詞だったものが、徐々に前置詞としての用法が強くなったものだが、賓語を取らない用法もあることから、必ずしも前置詞ともいえない、という厄介なものである。「堯以天下與舜」の「以」を動詞と見るならば、この文の *root* は「以」であり、「舜」は「與」の *obj* とみなされる。一方、「以」を前置詞と見るならば、この文の *root* は「與」であり、「天下」が「與」の *obj* であることを「以」が格表示 (*case*) しているのだから、「舜」は *iobj* とみなされる(図10)。われわれとしては、賓語を取らない「以」との兼ね合いで、あくまで「以」を動詞とみなしたいのだが、正直かなり悩ましい。

3.5 MISC

MISC には、現状では *SpaceAfter=No* だけを入れている。古典漢文においては、各単語の直後に空白などない、という現実をそのまま反映している。

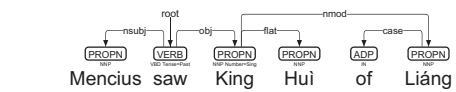
4. 他言語 UD との比較

われわれが構築中の古典中国語 UD は、はたして、UD として妥当性を持ったものなのか。その確信を少しでも得るべく、他言語 UD との比較を試みた。

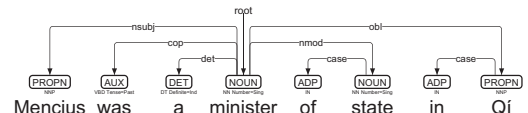
幸いなことに、Pulleyblank の古典中国語文法書 [4] では、各例文に英訳文が示されている。この英訳文に対し、英語 UD [7] を記述すれば、われわれの古典中国語 UD との比較が可能となるはずである。以下に、図2~10の各例文について、英語 UD との比較結果を述べる。

図2の「孟子見梁惠王」に対し、[4]の例文1が示している英訳は「Mencius saw King Hui of Liáng」である。この英訳文に対する UD 依存構造を、図11(a)に示す。「King」と「Hui」が *flat* で繋がれている点、および「of」が増えている点に違いはあるものの、ほぼ同じ UD 依存構造となっている。

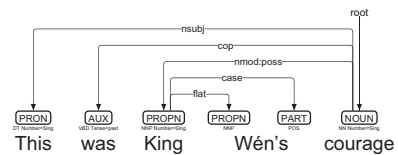
図3の「孟子為卿於齊」に対し、[4]の例文25が示している英訳は「Mencius was a minister of state in Qi」である。この英訳文に対する UD 依存構造を、図11(b)に示す。「卿」1字が「a minister of state」と英訳されている点については、



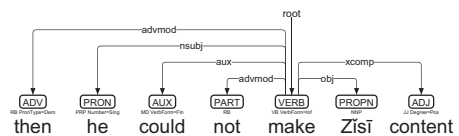
(a) 「Mencius saw King Hui of Liáng」の UD 依存構造



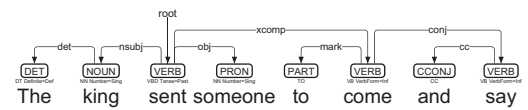
(b) 「Mencius was a minister of state in Qi」の UD 依存構造



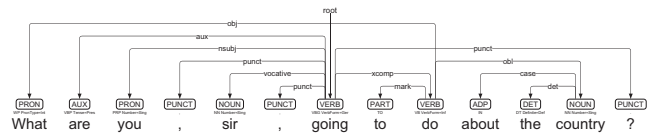
(c) 「This was King Wén's courage」の UD 依存構造



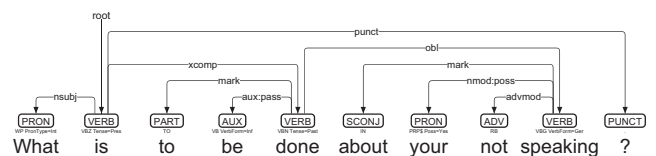
(d) 「then he could not make Zisi content」の UD 依存構造



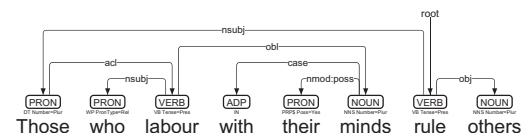
(e) 「The king sent someone to come and say」の UD 依存構造



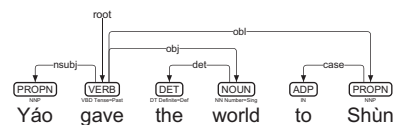
(f) 「What are you, sir, going to do about the country?」の UD 依存構造



(g) 「What is to be done about your not speaking?」の UD 依存構造



(h) 「Those who labour with their minds rule others」の UD 依存構造



(i) 「Yao gave the world to Shun」の UD 依存構造

図11 英語 UD との比較

*2 「者」の UD 品詞を、現在われわれは *PART* としている。これを *NOUN* にできるのなら *nsubj* はほぼ確実だし、*ADP* にできるのなら *mark* の可能性が上がる。しかしながら、古典中国語における「者」は、われわれの検討では「p, 助詞, 提示」であり、少なくとも *NOUN* とみなすのは難しい。

正直どうにもならないものの、その他の部分は、全く同じ UD 依存構造となっている。

図4の「此文王之勇也」に対し、[4]の例文12が示している英訳は「This was King Wén's courage」である。この英訳文に対する UD 依存構造を、図11(c)に示す。「was」が cop で繋がれている点、および「King」と「Wén」が flat で繋がれている点に違いはあるものの、ほぼ同じ UD 依存構造となっている。

図5の「則不能安子思」に対し、[4]の例文40が示している英訳は「then he could not make Zīsī content」である。この英訳文に対する UD 依存構造を、図11(d)に示す。「he」が主語として補われている上に、「安」を「make～content」と訳していることもあって、UD 依存構造がかなり異なってしまう。また、「not」が「make」に繋がれている点も、われわれの解釈とは異なる。

図6の「王使人來曰」に対し、[4]の例文109が示している英訳は「The king sent someone to come and say」である。この英訳文に対する UD 依存構造を、図11(e)に示す。「say」が「come」から conj で繋がれているという点を除けば、ほぼ同じ UD 依存構造となっている。

図7の「子若國何」に対し、[4]の例文86が示している英訳は「What are you, sir, going to do about the country?」である。図8の「若之何子之不言也」に対し、[4]の例文88が示している英訳は「What is to be done about your not speaking?」である。これらの英訳文に対する UD 依存構造を、図11の(f)と(g)にそれぞれ示す。「What」から始まる英語の UD 依存構造は、いずれも「若～何」とは全く異なっている。

図9の「勞心者治人」に対し、[4]の例文53が示している英訳は「Those who labour with their minds rule others」である。この英訳文に対する UD 依存構造を、図11(h)に示す。正直なところ、図9に示したわれわれの二案のいずれとも、あまり似ていない。しいて言えば、「者」を nsubj とみなした上の案だろうか。英訳文は「who」による従属節が長い上に、「labour」が目的語を取っておらず、その結果として UD 依存構造が、かなり異なってしまうのである(付録参照)。

図10の「堯以天下與舜」に対し、[4]の例文71が示している英訳は「Yáo gave the world to Shùn」である。この英訳文に対する UD 依存構造を、図11(i)に示す。図10に示したわれわれの二案のうち、下の案、すなわち「以」を前置詞とみなす案に、やや近い。ただし、この英訳文については、Pulleyblank 自身「Yáo with the world gave Shùn」を意識した結果だと述べており、さらなる検討が必要だろう。

5. おわりに

古典中国語に UD2.0 を適用すべく、UD 品詞と UD 依存構造について、具体的に検討をおこなった。UD 品詞につ

いては、われわれがこれまでにこなってきた MeCab 漢文形態素解析にもとづき、14の品詞への自動変換が可能であることを示した。一方、UD 依存構造については、Pulleyblank の古典中国語文法書の各例文について、手作業で UD 依存構造を決定していくことで、32種類の UD 依存構造タグが、古典中国語の構造記述に必要であることを明らかにした。UD2.0 に基づく古典中国語コーパスの作成を、本格的に開始できる準備が整った、と言えるだろう。

ただし、UD2.0 のもう一つの要素である UD 形態素属性については、われわれの MeCab 漢文形態素解析との齟齬が大きく、現時点では手も足も出ない。UD 形態素属性は、どちらかといえば屈折語における文法的機能を表すものが多く、孤立語である古典中国語には使えない。しかしながら、UD 形態素属性は、各言語ごとに提案が可能らしいので、今後はわれわれが使用している意味素性・小素性を、UD 形態素属性に含めていく方策を考えるべきだろう。

なお、本研究は、科学研究費補助金基盤研究(B)17H01835『古典漢文形態素コーパスにもとづく動詞の作用域の自動抽出』の研究助成を受けている。

付録 依存文法と UD 依存構造

UD 依存構造は、Мельчук の依存文法 [8] を理論的支柱としながらも、実践面では Reed-Kellogg の文法構造 [9] を取り入れている。これらについて、ざっと紹介しておこう。

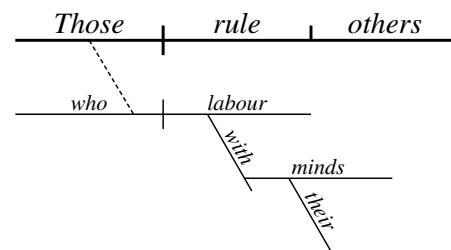


図12 「Those who labour with their minds rule others」の Reed-Kellogg 文法構造図

Reed-Kellogg 文法構造図は、主語 | 述語 | 目的語 という構造を基本に、英文を視覚化するものである。主文(main sentence)を太線で、それ以外の節(clause)を細線で視覚化した上に、修飾語を斜線に載せて示すという手法であり、たとえば「Those who labour with their minds rule others」という英文に対しては、「Those rule others」という主文の「Those」を、「who labour」という節が修飾していて、その「labour」を「with minds」が修飾し、「minds」を「their」が修飾していることから、図12のようになる。Reed-Kellogg 文法構造図は、英語に特化して設計されているが、現在も、アメリカの初等教育において幅広く用いられている。

Tesnière は、スラブ諸語の研究を通して、のちに依存文法と呼ばれる言語横断的な記述手法を提案した [10]。文に現れる語は互いに依存関係にある、というのが、依存文法

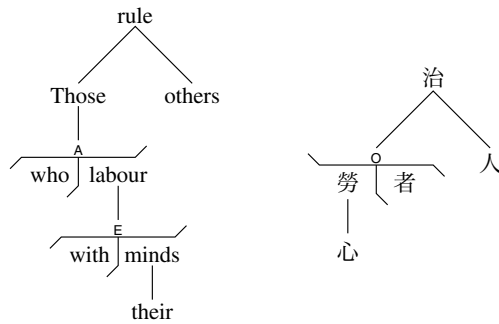


図13 「Those who labour with their minds rule others」(左)と「勞心者治人」(右)の Tesnière 依存文法図

の基本的な考え方である。語と語の間の依存関係においては、上位項 (régissant) が下位項 (subordonné) を支配し、下位項が上位項に依存する。たとえば「Those who labour with their minds rule others」という英文であれば、rule が Those と others を支配し、Those と others が rule に依存する。ただし、転用がおこなわれている場合は、それらの語の間には依存関係は無く、並置する形で表現する。たとえば図13左では、minds という名詞が with によって E 転用(連用修飾語に転用)されており、あるいは labour という動詞が who によって A 転用(連体修飾語に転用)されている。

Tesnière の依存文法は、文における語の順序に関わらず、語の依存関係を記述するものであり、たとえば古典中国語にも応用可能である。すなわち「勞心者治人」という文であれば、「心」が「勞」に依存し、「勞」が「者」によって O 転用(体言に転用)され「治」に依存する。これを図13右のように図示すれば、「Those who labour with their minds rule others」との間で、言語をまたいだ比較も可能となるのである。

Tesnière の依存文法は、その後、Robinson [11] や Hudson [12] らにより、Chomsky 句構造文法 [13] との融合が試みられた。その一方、Мельчук [8] は、語の依存関係の有向グラフ記述によって依存文法の形式化をおこない、Chomsky 句構造文法と決別した。

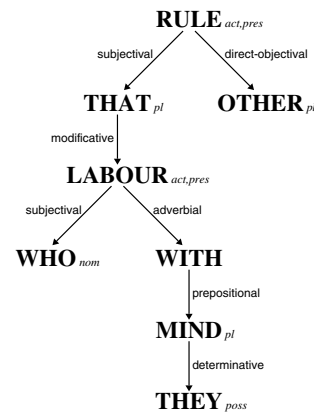
文中の語 X と Y に対し、Мельчук の依存文法は、 $X \rightarrow Y$ という依存関係によって文を記述する。 $X \rightarrow Y$ は、X は Y を支配する、あるいは、Y は X に依存する、と読む。各矢印には、依存関係に関する適切なラベルが付与される。各語が依存する語は高々1つであり、依存関係はループしない。依存関係は、あくまで語と語の間の記述であり、それがたとえ節 (clause) や句 (phrase) に係るものであっても、割り切って語と語の関係として記述する。

この「割り切り」の結果として、Мельчук の依存文法は、言語横断的な文法構造記述として、活用可能なものとなっている。英語の「Those who labour with their minds rule others」*3 と、日本語の「心を勞する者は人を治める」と、古典中国語の「勞心者治人」に対して、Мельчук の依存文法による文法構造記述を、図14に示す。ただし、ラベルのう

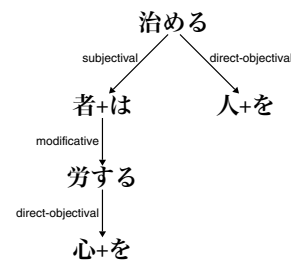
ち subjectival と direct-objectival は、過去には predicative と 1st completive が用いられており、Мельчук 自身も、ラベルがどれだけ必要なかは明らかにしていない [14]。

UD 依存構造は、Мельчук の依存文法におけるラベルを、表4の37種類に限定する野心的な試み、として捉えることができる。ただ、UD の前身である Stanford Typed Dependencies [15] が、初期には Bresnan 語彙機能文法 [16] からスタートしており、その後に依存文法へとシフトしたことから、いくつか不思議な概念が紛れ込んでしまっている。その一つが、節 (clause) である。

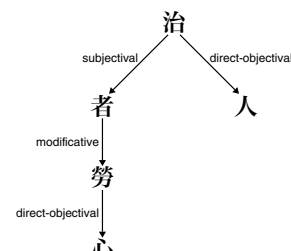
節という概念を、Мельчук は依存文法から削除したが、これを UD は再導入している。表4の nsubj と csubj は、いずれも主語を表す UD 依存構造タグであり、リンク先が節である時は csubj を、そうでない時は nsubj を使う。obj



(a) 「Those who labour with their minds rule others」*3



(b) 「心を勞する者は人を治める」



(c) 「勞心者治人」

図14 Мельчук の依存文法による文法構造記述

*3 図14(a)の WITH は、図13左の E 転用との兼ね合いを考えた場合、LABOUR ではなく MIND に依存すべきだという議論がある [17]。それを受けて、英語 UD では labour → minds → with となっている(図11(h))。英語の前置詞とそれに続く名詞類の扱いは、Tesnière と Мельчук と UD で、それぞれに異なる。

表4 UD 依存構造タグ

	Nominals	Clauses	Modifier Words	Function Words
Core arguments	nsubj obj iobj	csubj ccomp xcomp		
Non-core arguments	obl vocative expl dislocated	advcl	advmod discourse	aux cop mark
Nominal dependents	nmod appos nummod	acl	amod	det clif case
Coordination	MWE	Loose	Special	Other
conj cc	fixed flat compound	list parataxis	orphan goeswith reparandum	punct root dep

と ccomp についても同様だし, advmod と advcl についても同様だし, amod と acl についても同様である. case と mark については, リンク元が節である時は mark を, そうでない時は case を使う [18].

英語においては, 節という概念を導入するのは, そう難しくはない. たとえば, 図 11(h) の Those → labour に amod ではなく acl を付与しているのは, 端的には図 12 で「who labour」が節を構成している(というコンセンサスが得られている)からである. しかし, 英語以外の言語において, 節という概念は, 必ずしも自明ではない. そのせいもあって Мельчук は, 節や句といった概念を, 依存文法から削除したのである. これに対し, 依存文法に節を復活してしまった UD は, 言語横断を謳っているものの, 実は英語寄りになってしまっている可能性がある.

では, 古典中国語に対しては, UD 依存構造で導入されている節という概念は, 本質的に適用可能なのか. 端的には, 「勞心者」と「死者」の間には, 違いがあると考えべきなのか, そうでないのか. 残念ながら, 現時点のわれわれには, 確信も確証も得られていない. それもあって, 図 9 に示したわれわれの二案は, いずれとも決しがたく, また, 図 11(h) と単純に比較するわけにもいかないのである. 何とも悩ましい.

参考文献

[1] 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, 掲載予定).

[2] Daniel Zeman, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre, Filip Ginter, Juhani Luotolahti, Sampo Pyysalo, Slav Petrov, Martin Potthast, Francis Tyers, Elena Badmaeva, Memduh Gökırmak, Anna Nedoluzhko, Silvie Cinková, Jan Hajič jr., Jaroslava Hlaváčová, Václava Kettnerová, Zdeňka Uřešová, Jenna Kanerva, Stina Ojala, Anna Missilä, Christopher Manning, Sebastian Schuster, Siva Reddy, Dima Taji, Nizar Habash, Herman Leung, Marie-Catherine de Marneffe, Manuela Sanguinetti, Maria Simi, Hiroshi Kanayama, Valeria de Paiva, Kira Droganova, Héctor Martínez Alonso, Çağrı Çöltekin, Umut Sulubacak, Hans

Uszkoreit, Vivien Macketanz, Aljoscha Burchardt, Kim Harris, Katrin Marheinecke, Georg Rehm, Tolga Kayadelen, Mohammed Attia, Ali Elkahky, Zhuoran Yu, Emily Pitler, Saran Lertpradit, Michael Mandl, Jesse Kirchner, Hector Fernandez Alcalde, Jana Strnadová, Esha Banerjee, Ruli Manurung, Antonio Stella, Atsuko Shimada, Sookyoung Kwak, Gustavo Mendonça, Tatiana Lando, Rattima Nitisaroj, and Josie Li: Multilingual Parsing from Raw Text to Universal Dependencies, Proceedings of CoNLL 2017 Shared Task (August 2017), pp.1-19.

[3] Daniel Andor, Chris Alberti, David Weiss, Aliaksei Severyn, Alessandro Presta, Kuzman Ganchev, Slav Petrov and Michael Collins: Globally Normalized Transition-Based Neural Networks, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (August 2016), pp.2442-2452.

[4] Edwin G. Pulleyblank: Outline of Classical Chinese Grammar, Vancouver: UBC Press (1995).

[5] Herman Leung, Rafaël Poirer, Tak-sum Wong, Xinying Chen, Kim Gerdes and John Lee: Developing Universal Dependencies for Mandarin Chinese, 12th Workshop on Asian Language Resources (December 12, 2016), pp.20-29.

[6] 山崎直樹, 守岡知彦, 安岡孝一: 古典中国語形態素解析のための品詞体系再構築, 人文科学とコンピュータシンポジウム「じんもんこん 2012」論文集(2012年11月), pp.39-46.

[7] Natalia Silveira, Timothy Dozat, Marie-Catherine de Marneffe, Samuel R. Bowman, Miriam Connor, John Bauer and Christopher D. Manning: A Gold Standard Dependency Corpus for English, Proceedings of the 9th International Conference on Language Resources and Evaluation (May 2014), pp.2897-2904.

[8] Igor A. Mel'čuk: Dependency Syntax: Theory and Practice, New York: State University of New York Press (1988).

[9] Alonzo Reed and Brainerd Kellogg: Higher Lessons in English: A Work on English Grammar and Composition, New York: Clark & Maynard (1877).

[10] Lucien Tesnière: Éléments de Syntaxe Structurale, Paris: C. Klincksieck (1959).

[11] Jane J. Robinson: Dependency Structures and Transformational Rules, Language, Vol.46, No.2 (June 1970), pp.259-285.

[12] Richard Hudson: Word Grammar, New York: Basil Blackwell (1984).

[13] Noam Chomsky: Aspects of the Theory of Syntax, Cambridge: MIT Press (1965).

[14] Alain Polguère, Igor A. Mel'čuk: Dependency in Linguistic Description, Amsterdam: John Benjamins (2009).

[15] Marie-Catherine de Marneffe and Christopher D. Manning: The Stanford Typed Dependencies Representation, Coling 2008: Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation (August 2008), pp.1-8.

[16] Joan Bresnan: Lexical-Functional Syntax, Malden: Blackwell (2001).

[17] 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.

[18] Marie-Catherine de Marneffe, Timothy Dozat, Natalia Silveira, Katri Haverinen, Filip Ginter, Joakim Nivre, Christopher D. Manning: Universal Stanford Dependencies: A Cross-Linguistic Typology, Proceedings of the 9th International Conference on Language Resources and Evaluation (May 2014), pp.4585-4592.