

「単位文脈構造体」言語モデル

漆原憲博¹¹株式会社ジェーエフピー japanfp@jfp.co.jp

概要

ソフトウェア要求文書やまた特許文書など多くの条件文からなる文書である。読み手は条件が複雑に入り組んだとき理解に手こずる。条件に依存する文書は、通常の業務文書にも当てはまり、手強いところでは国際標準規格文書にも当てはまる。本稿では条件文の多い文章に対して、理解精度の向上と時間の節約を促進する策を講じる。加えて、第三者による知見も活かしたい。これらの課題の解決策として言語モデル「単位文脈構造体」を提案する。

1 はじめに一課題—

自動運転や医療系の安全やサイバーセキュリティのシステムの開発業務においては、国際標準規格 (ISO, IEC) や国権規格 (UN) の文書を理解することが求められる。開発システムがこれら文書の規格 (以下、規格と呼ぶ。) に適合しなければならないからである。ここでシステムとは、自動制御機器であったり、同機器の作動する作業現場全体であったりする。作業現場には人間も含まれる。このようなシステムに対して、規格の実現が求められる。

このような中、開発業務に携わる、例えばシステム設計者は、規格の理解に困難を覚える。規格の理解者は通常専門の担当者が存在する。よって、開発者は専門の担当者に支援を求めればよい。しかし、ときに専門担当者の理解が必ずしも正確でない場合がある。さらに理解の支援にはコンサルタントも存在する。コンサルタントは様々な企業の事情に詳しく、また規格の変更にも通曉している。しかし、理解が正確であるという保証はない。

以上のことから、開発者は理解が腑に落ちない場合も少なくない。規格の理解者たちが、開発者の開発する対象を理解していない場合などは特にそうである。その結果、開発者が規格文書それ自体を確認するために多くの時間を費やすことも少なくない。

しかし、規格文書は条件が多く複雑で分厚く、他文書との関係も多いため理解が難しい。分厚いため、条件を最後まで徹底して記述しきれず、途中で

中に条件を代替する注釈がある。これも条件を増やす。他文書との関係は、規格の親、子、また兄弟という関係がある。親が子の概念を規定する。

また、規格には開発対象システムの規格遵守に関する審査制度を設けているものもある。

以上が開発者を取り巻く規格の事情である。といって、開発者はこれらの規格すべてを理解する必要はない。開発者はシステムの要求事項を決めることが役割である。この限りにおいて規格を理解すればよい。ただし、例えば自らの開発内容に近い兄弟規格の理解は、自らのシステムの要求事項の作成に大いに参考になる。このことから、やはり開発者は多くの時間を規格の理解のために割くといえる。

多くの時間を要することは、専門担当者やコンサルタントにも当てはまる。審査制度を預かる審査官もまた然りである。

さらに、一つの文書にこのように多くの立場の人たちが関与するという事は、それぞれの立場からの理解の違い、いわば知見が必要事項として加わる。

理解の時間の節約と理解精度の向上に加え、必要な知見を、対象となる文書群に付加できるとよい。対象となる文書群は条件の多いことに特徴を持つ。開発者、専門担当者、コンサルタント、審査官に効用が及ぶ。そのための良い方法はないか、これが本稿の課題である。

2 課題の解決案

「単位文脈構造体」(以下、文脈構造体と略す場合もある。) という言語モデルにより先の課題を解決する。文脈構造体は任意の文書を条件や連言からなる文の連なり、すなわち文脈として再構成するものである。これにより、文書の理解がより短い時間で、精度高く行われ、かつ文書に各知見が加わる事が期待される。

図1に文脈構造体の位置づけを示す。図1では、既存の文書を解析し、文脈構造体を作成し、文脈構造体に対して更新(修正、削除、追加)し、新文書となることを示している。



図 1 □単位文脈構造体の位置づけ

2.1 単位文脈構造体の構成

単位文脈構造体は以下の構成要素からなる。

- (1) 単位文：文の最小単位を単位文とする。単位文は、対象を指す対象語句と、対象の属性を表現する属性語句からなる。例えば、「安全は危険から自由である」という文は、例えば括弧付きで再構成され、単位文「<安全>は {危険から自由である}」で表現される。格助詞「は」は無くてもよい。
- (2) 接続関係語句：単位文は、「ならば」、「である場合」や「かつ」、「または」などの接続関係語句により、単位文どうしが結ばれるとする。
- (3) 文脈：結ばれ方は、単位文が「ならば」で結ばれた後に、次にすぐに「かつ」や「または」で複数個結ばれることも普通なので、その結びの様子を称して「文脈」と呼ぶこととする。文脈は「ならば」などの条件文を含む。
- (4) 単位文の結合子：単位文を繋ぐ結合子は論理学でいう含意や連言とする。選言（「または」）は連言と否定で表現し得るので、文脈構造体では連言を用いることを優先する。経験的に、選言以上に連言の理解の方が理解に要する時間が短いと思われるからである（ただし、検証が要る）。
- (5) 否定前件含意：「ならば」($P \Rightarrow$)という含意の接続関係語句に対しては、必ず「そうでないならば」($\neg P \Rightarrow$)という前件 P の否定を前件とする含意を、文脈構造体に導入する。

2.1.1 否定前件含意の必要性

文脈構造体においては否定前件含意を必須とする。条件の網羅性確保と矛盾回避のためである。

チョムスキーの有名な例文”Colorless green ideas sleep furiously.”が、この網羅性のヒントである。

この意味不明な文を有意味な文にしようとするとき、いろいろなケースを考えようと思いが働く。色が無いのに緑とはおかしい、また緑を主において、緑なのに色がないというのも変である、などと考える。双方の概念は矛盾している。この矛盾の解消のために、例えば「もし色が無いというのなら、緑と

矛盾しないように、緑から色を取り払うような、架空の設定を考える」というように思考は進む。他方、色のある緑を前提にすれば、「色の無い」という表現を「実は目の見えない人が、, , 」などと架空の場面を設定する。

以上のように、我々は矛盾を解消するために、文の意味を反対にした、無矛盾の状況を設定することを試みる。命題「 P ならば」に対して、その反対「 $\neg P$ ならば」を設定し、全体を無矛盾になるように図る。

「 $\neg P$ ならば」を事例で示せば、後掲の表 2 の 07 行目の「でないとき」という表現である。これは 04 行目の「もし<速度制御>が {適切である}ならば」（「 P 」）に対応している。つまり、 P を否定している。

このように含意を相補的に用いて、意味対象が矛盾に陥ることを防いでいる。これを相補的含意 (Complementary Implication) と呼ぶこととする。この相補的含意を用いることで意味対象の全領域 (Universe of Discourse) をカバーすることを図る。

この適用は、あからさまに矛盾ではない意味の対象、すなわち通常の文書に対しても適用できる。つまり、時に文書は、規約の条件が間違っていたりする。また、想定外の事柄も発生する。

そこで、文書の中に間違いがあるかも知れないという前提で、修正の余地を残すために「否定前件含意」を文書に相当する文脈構造体に組み込む。これにより、文書が矛盾や条件設定の不足から逃れることができる。

文脈構造体は、否定前件含意のみならず、新たに条件を追加できる機能も持つことにする。これにより、先述の各立場の知見も反映できる。

文脈構造体の相補的含意による網羅的構造は、チョムスキーの「普遍文法」ならぬ、「普遍文脈」という新たな用語を連想させる。

2.2 単位文脈構造体の応用事例

単位文脈構造体の応用事例を示す。先の課題からいえば、規格の事例がよい。が、規格は条件文が多いことが特徴である。そこで事例には、条件文を明示的に書いている文章を選ぶ。

特許文は好例である。これを規格に代替する。

表 1 文書：特許文章

この縦位置の誤差が大きいと、自動車が直線路からカーブ路に進入した場合、車両の速度が大きすぎるときでも、もし速度制御が適切であるならば、自動車はうまく曲がり切れ、カーブ路の走行車線内の走行を自動でできる。（引用：特許特開 2018-30495）

この例の文脈構造体は以下となる。条件を表わす記号として「もし」、「でないとき」、「おわり」を使っている。他の記号でもよい。ただし、「もし」～「おわり」は1つの組である。文脈構造体では組として存在する必要がある。「もし」が続く場合には「もし」～「おわり」は入れ子構造になる。

表 2 単位文構造体例

01	もしこの縦位置の誤差が {大きい} と、
02	もし<自動車>が {直線路からカーブ路に進入した} 場合、
03	もし<車両の速度>が {大きすぎる} ときでも、
04	もし<速度制御>が {適切である} ならば、
05	そのときには<自動車>は {うまく曲がり切れ}、
06	かつ <カーブ路の走行車線内の走行>を {自動でできる}
07	でないとき
08	おわり
09	でないとき
10	おわり
11	でないとき
12	おわり
13	でないとき
14	おわり

次に表 3 は 08 行目（下線）では、04 行目の条件文「もし、」の否定「でないとき、」（07 行目）の次の行 08 行目に人為的に単位文を追記している。文脈構造体は反対条件の有無を常に問う（作る）ことができる。これにより条件の抜けがないか確認できる。08 行以降も単位文を追加している。

表 3 単位文構造体の更新

01	もしこの縦位置の誤差が {大きい} と、
02	もし<自動車>が {直線路からカーブ路に進入した} 場合、
03	もし<車両の速度>が {大きすぎる} ときでも、
04	もし<速度制御>が {適切である} ならば、
05	そのときには<自動車>は {うまく曲がり切れ}、
06	かつ <カーブ路の走行車線内の走行>を {自動でできる}
07	でないとき
08	そのときには<自動車>は {うまく曲がり切れる}、
09	かつ <カーブ路の走行車線内の走行>を {自動でできる}
10	おわり
11	でないとき
12	そのときには<自動車>は {うまく曲がり切れ}、
13	かつ <カーブ路の走行車線内の走行>を {自動でできる}
14	おわり

15	でないとき
16	そのときには<直線路の走行車線内の走行>を {自動でできる}
17	おわり
18	でないとき
19	もし<自動車>が {直線路からカーブ路に進入した} 場合、
20	もし<車両の速度>が {大きすぎる} ときでも、
21	もし<速度制御>が {適切である} ならば、
22	そのときには<自動車>は {うまく曲がり切れ}、
23	かつ <カーブ路の走行車線内の走行>を {自動でできる}
24	でないとき
25	そのときには<自動車>は {うまく曲がり切れ}、
26	かつ <カーブ路の走行車線内の走行>を {自動でできる}
27	おわり
28	でないとき
29	そのときには<自動車>は {うまく曲がり切れ}、
30	かつ <カーブ路の走行車線内の走行>を {自動でできる}
31	おわり
32	でないとき
33	そのときには <直線路の走行車線内の走行>を {自動でできる}
34	おわり
35	おわり

文脈構造体は文脈（条件等）を個別に扱うことができる。よって、文脈の比較が容易である。例えば、2つの文脈において、条件の前件部が同じで、かつ後件部が異なるならば、一方の文脈が間違っている可能性がある。

この比較において、対象語句と属性語句の区別は、比較の機械的実行を容易にする。対象語句の持つてはならない属性も事前に設定できる。

2.2.1 文脈の明示

表 3 の 01～06 を連言で結び、含意文の前件部となし、05 と 06 行を連言で結び、後件部とすると1つの文脈が明示的になる。また、01～03 と 07～09 は別な文脈を作っている。表 4 にこれら 2 例の文脈をあげる。読み手は文脈単位に文を理解できる。

表 4 文脈単位の切り出し

[文脈 1]
<この縦位置の誤差>が {大きい}
∧<自動車>が {直線路からカーブ路に進入した}
∧<車両の速度>が {大きすぎる}
∧<速度制御>が {適切である}
⇒<自動車>は {うまく曲がり切れ}
∧<カーブ路の走行車線内の走行>を {自動でできる}
[文脈 2]
<この縦位置の誤差>が {大きい}
∧<自動車>が {直線路からカーブ路に進入した}
∧<車両の速度>が {大きすぎる}
∧<速度制御>が {適切である} = {適切でない}
⇒<自動車>は {うまく曲がり切れる}、

△<カーブ路の走行車線内の走行>を{―自動でできる}

2.2.2 新文書の作成

更新された単位文構造体からは、新たな文書を作成できる。表5は、表3から生成AIにより出力された新文書である。自然文である。文脈構造体自体を記述できる。出力された自然文は網羅性を持つ。

表5 新文書の生成

この縦位置の誤差が大きい場合、自動車が直線路からカーブ路に進入したとき、車両の速度が大きすぎる場合でも、速度制御が適切であれば、自動車はうまく曲がり切れ、カーブ路の走行車線内を自動で走行することができる。

しかし、速度制御が適切でない場合には、自動車はうまく曲がり切れず、カーブ路の走行車線内を自動で走行することはできない。

また、車両の速度が大きすぎる状況でない場合には、自動車はうまく曲がり切れ、カーブ路の走行車線内を自動で走行することができる。

さらに、自動車が直線路からカーブ路に進入していない場合には、直線路の走行車線内を自動で走行することができる。

加えて、この縦位置の誤差が大きくない場合にも、自動車が直線路からカーブ路に進入したとき、車両の速度が大きすぎる場合でも、速度制御が適切であれば、自動車はうまく曲がり切れ、カーブ路の走行車線内を自動で走行することができる。

しかし、速度制御が適切でない場合には、自動車はうまく曲がり切れず、カーブ路の走行車線内を自動で走行することはできない。

また、車両の速度が大きすぎる状況でない場合には、自動車はうまく曲がり切れ、カーブ路の走行車線内を自動で走行することができる。

さらに、自動車が直線路からカーブ路に進入していない場合には、直線路の走行車線内を自動で走行することができる。

2.3 課題：文の接続関係

自然文には文の接続関係を表わす接続関係語句が明示的でないものもある。さらに考察が要る。

また、条件に相当する表現として、どの文書であれ、項番を多く使う。もっとも項番なかに条件が明示的に書かれているのも多い。他方、項番のみで条件の違いを表現しているのも普通である。

注釈も条件を示す。注釈の条件が、記述箇所の項に属するのであれば、文脈構造体に反映することは容易である。文書全体に関わる注釈の場合、前提条件の設定位置には要注意である。文書全体の第一前提におくべき場合もある。

行と列からなる表を条件文として、文の連続からなる文脈に移し代えることも、時には有効である。表は一覧性のよさから、ときに厳密な場合分けの網

羅性を見逃すことがあるからである。文脈構造体の作成にはこれらを考慮する必要がある。

疑問文は後続の文に繋がる接続関係語句といえる。

3 効果

文書の読み手は、表4のような文脈ごとに明確に場合分けされた知識を学ぶことで、知識の精度を上げ、時間も節約できる。例えば「どんな場合にクルマはカーブに注意しなければならないか」と問うと文脈構造体は、表4のような明確な文脈をベースに回答をする。回答が導出される過程も明確である。

また極端だが、読み手は、知識を事前に学ぶことなく、都度問うことで、文脈構造体は周到に回答する。文脈構造体を使うツールを機能アップできよう。

文脈構造体は複数の文書を統合できる。条件の同じ文脈を統合することで知識をまとめることができる。当然文書ごとの区別もできる。時間の節約と理解の精度の向上に役立つ。そう期待できる。

知識の記憶教育に多くの時間が割かれている学校教育の改善にも期待できる。記憶教育から考える教育へ多くの時間を配分できる。

4 おわりに―応用、発展、研究―

今後の実用化に向けては、国際標準規格文書、および広く業務規約・契約文書への適用試験が要る。

なお、単位文を基本とする文の構成は、特定の言語を超える。動物言語学者の鈴木俊貴氏ではないが、単位文は氏の唱えるシジューカラの言語と同じく対象の指示とその属性からなる。言語活動のベースと推測される。とすれば、本稿の構想は極めて基礎がかたいと推量される。応用範囲の拡大が期待できる。

文脈構造体は言語モデル(LM)の一種である。多くの文書を読込み大規模化することも可能である。文は含意により分岐されるので、知識の導出過程が明確である。

また、文脈構造体における基本語句は対象語句と属性語句である。かつ、これらの語句は文脈に挟まれて在る。このことから、注力(attention)をこれら基本語句と文脈の、いわば挟み構造に対して行う計算効率化の、電力消費に関わる研究課題が生まれる。

5 参考文献

本稿は『「文脈ベース」、文関連装置』(特許7706687)を参考にしている。