組合せ範疇文法を用いたドイツ語文の漸進的解析

高橋直人　竹内泉　一杉裕志
産業技術総合研究所人工知能研究センター
{naoto.takahashi,takeuti.i,y-ichisugi}@aist.go.jp

概要

文法的には異なる統語構造を持っているにもかかわらず，人間にとっては同じ意味と理解される複数の文があった場合に，それぞれの文を「漸進的に」解析し，かつ同一の意味表示を得る事が可能か否かを検討する．解析対象としては，接続語のタイプによって語順がさまざまに変化するドイツ語文を用いる．また文法形式としては，漸進的解析に適しているとされる組合せ範疇文法を用いる．

1　はじめに

人間が自然言語を理解する際は，文の終了を待たず，文頭から漸進的に解析を行っていると考えられる．またその推測を支持する心理学的実験結果も存在する [1, 2]．更に工学的な応用として同時通訳や字幕生成などの自動化を考慮した場合，漸進的な解析技術は不可欠であると言える．

一概に文頭からの漸進的解析と言っても，そこにはさまざまなレベルがありうる．比較的制限の緩い方法としては，語順に従って 1 語ずつ読み進めながら作成した複数の部分木を，それぞれ個別の記憶領域に一時的に保存し，他の部分木との接続関係が明確になった時点で初めて互いに連結する，という戦略が考えられる．一方，厳密な漸進的解析としては，文頭から順に 1 語読み進むたびに，その語を現在作成中の部分木の一部として即座に埋め込む，という方式がある．

本稿では，人間の文理解モデルに関して考察する．具体的には，統語構造的にはさまざまな形を取るが，意味内容的には同一の複数のドイツ語文を，組合せ範疇文法を用いて「厳密な形で」漸進的に解析する際に，どのような語彙項目と導出過程が必要となるか考察する[1)]．

1）　人間が厳密な漸進的解析を行っていると主張しているわけではなく，一つの可能性としての考察である．

順関数適用規則　$\dfrac{X/Y:f \quad Y:a}{X:fa}$ >

逆関数適用規則　$\dfrac{Y:a \quad X\backslash Y:f}{X:fa}$ <

順関数合成規則　$\dfrac{X/Y:f \quad Y/Z:g}{X/Z:\lambda x.f(gx)}$ >B

$\dfrac{X/Y:f \quad (Y/W)/Z:g}{(X/W)/Z:\lambda x.\lambda y.f(gxy)}$ >B^2

順型繰り上げ規則　$\dfrac{X:a}{T/(T\backslash X):\lambda f.fa}$ >T

図 1　本稿で使用する組合せ規則一覧

2　組合せ範疇文法

組合せ範疇文法 [3, 4, 5, 6] は語彙化文法の一種であり，「マイルドな文脈依存」と言われる弱生成能力を持つ．そのため，文脈自由文法の範囲に収まらない自然言語現象を記述する事ができる．また理論的にも実際的にも効率のよい解析アルゴリズムが知られている [7, 8] ため，自然言語の文法記述に適しているとされている．

更に，統語構造から意味表示への直接的なマッピングがあるので，複数の文を統語解析した後，それらの意味表示を比較するのに都合が良い．

組合せ範疇文法では，S（文）や NP（名詞句）のような少数の基底範疇と，それらを結ぶ演算子を再帰的に用いる事で，多様な統語範疇を表現する．演算子には / と \ がある．X と Y を統語範疇としたとき，X/Y は「右側に Y が隣接したときに X となる統語範疇」，$X\backslash Y$ は「左側に Y が隣接したときに X となる統語範疇」を意味する．

このように作られる統語範疇の構造は非常に柔軟であり，未完成の統語的構成素を自然に表現することができる．それゆえに組合せ範疇文法は，他の文法形式に比べて漸進的な構文解析に向いているとさ

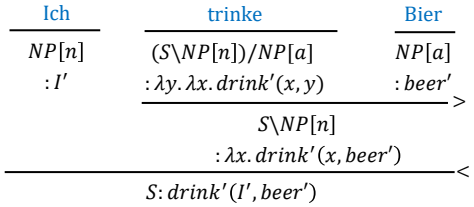

図 2　“Ich trinke Bier” の通常の（非漸進的）導出．S はすべて $S[dcl]$ すなわち平叙文．

れている [1, 2, 3, 9]．

たとえば “The boy will eat the cake” という文があった時，厳密な漸進的解析を行うためには “The boy will eat the” に対応する部分にも統語範疇を割り当てる必要がある．一般の文脈自由文法でそのような統語範疇を設定するのは困難であるが，組合せ範疇文法であれば S/N，つまり「右に N（名詞）が隣接した場合に S（文）となる範疇」と自然に表現することができる [1]．

語彙項目に付与された統語範疇は，少数かつ言語非依存の組合せ規則によって互いに組み合わされるが，このとき意味表示も同時並列的に合成される．この操作が再帰的に実行される事によって，語彙項目レベルの意味表示が文レベルの意味表示へと順次成長する．

本稿で使用する組合せ規則の一覧を図 1 に示す．コロンの左側が統語範疇で，右側がラムダ項による意味表示である．

図 2 にドイツ語文 “Ich trinke Bier”（英訳 “I drink beer”）の通常の非漸進的導出を，図 3 に同じ文の漸進的導出を示す．多くの場合は図 3 のように，型繰り上げ規則と関数合成規則を組み合わせる事で漸進的な導出が可能になるが，標準的な組合せ規則だけでは漸進的に導出できないパターンも存在する [1]．

3　漸進的ドイツ語解析

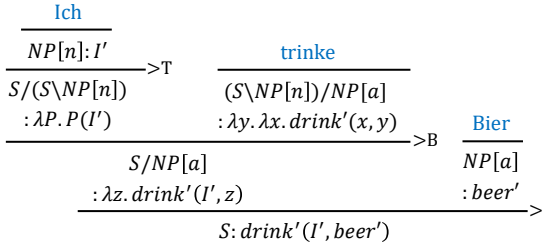

図 3　“Ich trinke Bier” の漸進的導出．S はすべて $S[dcl]$．

1. 従属複合文（原因文の文頭に従属接続詞）
Ich trinke Bier, weil ich Durst habe.
I drink beer because I thirst have

2. 従属複合文（1 の変形；原因文が先行）
Weil ich Durst habe, trinke ich Bier.
because I thirst have drink I beer

3. 並列複合文（原因文の文頭に並列接続詞）
Ich trinke Bier, denn ich habe Durst.
I drink beer because I have thirst

4. 並列複合文（原因文の文中に接続的副詞）
Ich trinke Bier, ich habe nämlich Durst.
I drink beer I have because thirst

5. 並列複合文（結果文の文頭に接続的副詞）
Ich habe Durst, deshalb trinke ich Bier.
I have thirst therefore drink I beer

6. 並列複合文（結果文の文中に接続的副詞）
Ich habe Durst, ich trinke deshalb Bier.
I have thirst I drink therefore beer

図 4　「喉が渇いたのでビールを飲む」のドイツ語訳 6 種

3.1　原因・結果関係の統語的パターン

他の自然言語同様，ドイツ語の原因・結果関係は，図 4 に示すようにさまざまな統語構造で表現される．たとえば文 1，3，4 を比較すると，先行文が全く同一であっても，接続関係を示す語（weil, denn, nämlich）の違いによって，後続文の語順が変化する事が分かる．したがって，接続関係を示すこれらの語には，それぞれ異なった統語範疇を付与する必要がある．

このような統語構造とその意味表示の複雑な関係を漸進的に解決する場合に，どのような語彙項目と導出過程が考えられるかを以下に示す．

3.2　語彙項目と導出過程

本稿では，基底範疇 NP（名詞句）の統語素性に関して n（主格）と a（対格）の 2 種類を区別する．また基底範疇 S（文）の統語素性に関しては，Hockenmaier[10] を踏襲し，*dcl*（declarative，定動詞第 2 位），*v1*（verb initial，定動詞文頭），*vlast*（verb final，定動詞文末），*emb*（embedded，副文）の 4 種類を使用する．ドイツ語動詞の統語範疇は，英語動詞同様に基底型ではなく関数型とするが，文頭の定動詞と文頭以外の定動詞で意味表示が異なる点に注意が必要である．

また，weil, denn, nämlich, deshalb 等の接続詞および接続的副詞に関する語彙項目は，一般的な英語の接続詞や副詞と異なる部分が多い．これは原因・結果関係を示すドイツ語固有の統語構造と，本稿の目

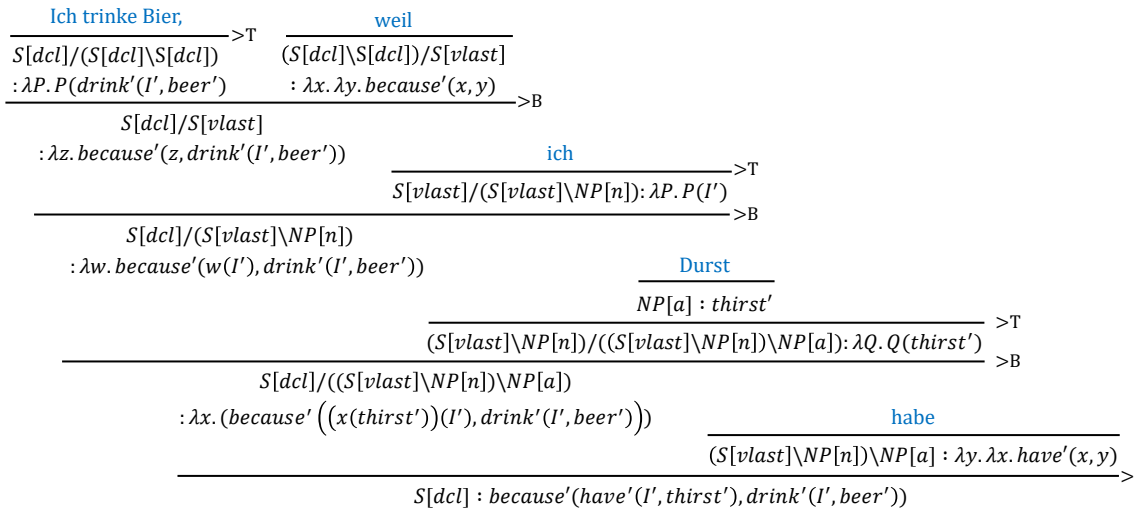

図 5　“Ich trinke Bier, weil ich Durst habe.” の漸進的導出

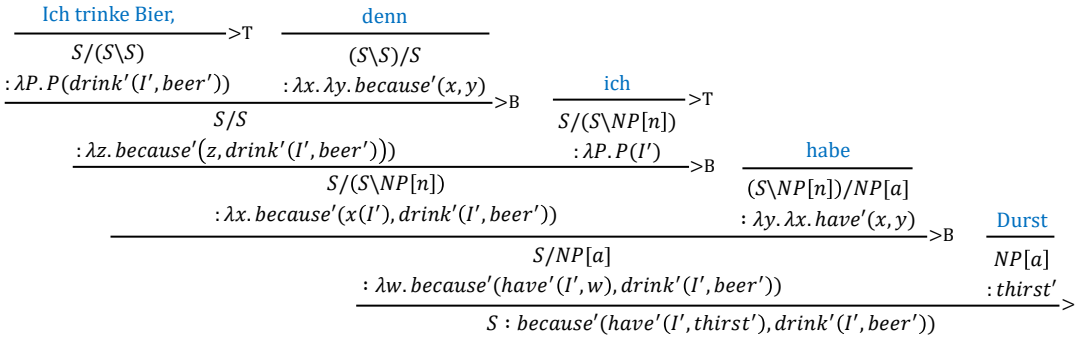

図 6　“Ich trinke Bier, denn ich habe Durst.” の漸進的導出．S はすべて $S[dcl]$．

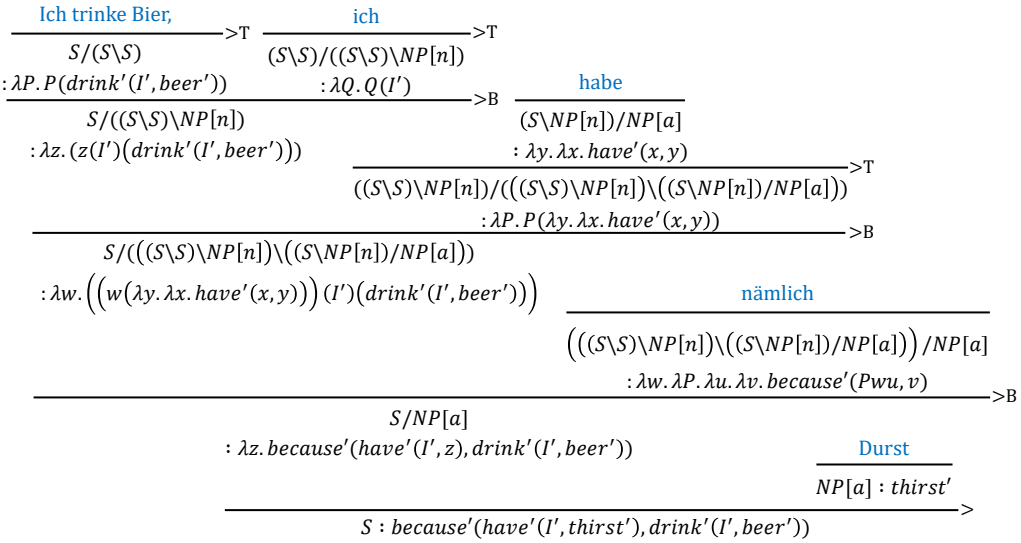

図 7　“Ich trinke Bier, ich habe nämlich Durst.” の漸進的導出．S はすべて $S[dcl]$．

的である厳密な漸進的解析を実現するためである．

以上を踏まえ，厳密な漸進的解析が可能になるように設定した語彙項目と，そのときの各文の導出過程（の一例）を図 5 から図 10 に示す．最終的にどの文からも同一の意味表示が得られている事が分かる．

4 関連研究

Stanojević and Steedman [2] は，漸進的木回転（Incremental Tree Rotation）と名付けた組合せ範疇文法の解析アルゴリズムを提案している．このアルゴリズムは，右枝分かれの木を左枝分かれに変換すると共に，右端に出現する可能性を秘めた潜在的な修飾語句に対応した漸進的な処理が可能である．漸進的木回転は純粋に文法的な手法であるが，意味表示やヒューリスティックスを利用する先行研究に比べて，より高い漸進性と精度を示すと報告されている．

Demberg [1] は組合せ範疇文法における厳密な漸進性（strict incrementality あるいは full connectedness）について議論し，どのような場合に厳密な漸進的解析が可能で，どのような場合に不可能であるかを分析している．また一般の組合せ範疇文法には陽に含まれない「Geach 規則」を導入し，厳密な漸近的解析が適用できる範囲を拡大する方法を示している．

$$\text{Geach 規則 }: Y/Z \Rightarrow_{\mathrm{B}} (Y/G)/(Z/G)$$

語彙項目としては通常の非漸進的解析に用いられる物をそのまま変更せずに採用しつつ，組合せ規則を追加することで漸近的に解析可能な範囲を拡張している点が本稿との違いである．

ブログ記事 [11, 12] は，上記 Demberg [1] を踏まえた上で一般化型繰上げ規則と一般化 Geach 規則を提案し，更に以下のような予想を与えている．

> application/composition/type raising のみが許される CCG の任意の導出木を，generalized type raising と generalized Geach rule を使って，（semantics を変えることなく）incremental な導出木に変形できることが示せると思う

興味深い予想であり，今後の展開が期待される．

5 おわりに

意味内容的には同一だが，統語構造的にはさまざまな形を取る複数の文を，組合せ範疇文法を用いて厳密な形で漸進的に解析しつつ，最終的に同一の意味表示を得る事ができるか否かを，ドイツ語の原因・結果関係文を題材として考察した．

本稿の漸進的解析では，型繰り上げ規則と関数合成規則を使用しているが，これらの規則を多用すると擬似的曖昧性（spurious ambiguity）が急激に増加してしまう，という問題が知られている．そのため正規型（normal form）と呼ばれる導出においては，関数合成と型繰り上げの使用は関係節など統語的に必要な場面に制限されている [9] .

また，Hockenmaier and Steedman[13] においても，関数合成と型繰り上げは必要な場合のみ行うとされており，更に Stanojević and Steedman[2] では，型繰り上げを適用する範疇は *NP* や *PP* などの引数に限られる，とされている．

異なる統語構造に対して厳密な漸進的解析を実行し，最終的に同一の意味表示を得るという当初の目的は達成されたが，厳密な漸進性を重要視した結果として逆に探索空間が増加してしまうのでは，人間の言語理解モデルとして十分ではないと考えている．また図 7 の nämlich や図 10 の deshalb に付与した語彙項目は，厳密な漸進性を実現するためとは言え，複雑すぎるようにも見える．

今後は漸進性のレベルと探索空間の範囲，それに語彙項目の複雑さのバランスが取れた言語理解モデルの構築に取り組みたい．

謝辞

本研究は JSPS 科研費 JP22K12188 の助成を受けたものです.

参考文献

[1] Vera Demberg. Incremental derivations in CCG. In **Proceedings of the 11th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+11)**, pp. 198–206, Paris, France, September 2012.

[2] Miloš Stanojević and Mark Steedman. CCG parsing algorithm with incremental tree rotation. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 228–239, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[3] Mark Steedman. **The Syntactic Process**. The MIT Press, 2000.

[4] Mark Steedman. **Surface Structure and Interpretation**. The MIT Press, 1996.

[5] 戸次大介. 日本語文法の形式理論. くろしお出版, 2010.

[6] 峯島宏次. 論理と文法. 数学セミナー, Vol. 59, No. 3, pp. 30 – 35, 2020.

[7] Krishnamurti Vijay-Shanker and David Weir. Polynomial time parsing of combinatory categorial grammars. In **28th Annual Meeting of the Association for Computational Linguistics**, pp. 1–8, 1990.

[8] Stephen Clark and James R. Curran. Wide-coverage efficient statistical parsing with CCG and log-linear models. **Computational Linguistics**, Vol. 33, No. 4, pp. 493–552, 2007.

[9] David Reitter, Julia Hockenmaier, and Frank Keller. Priming effects in Combinatory Categorial Grammar. In **Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing**, pp. 308–316, Sydney, Australia, July 2006. Association for Computational Linguistics.

[10] Julia Hockenmaier. Creating a CCGbank and a wide-coverage CCG lexicon for German. In **Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics**, pp. 505–512, Sydney, Australia, July 2006. Association for Computational Linguistics.

[11] 氏名不詳. 組合せ範疇文法の漸進的構文解析, 2021. https://m-a-o.hatenablog.com/entry/2021/05/05/085810.

[12] 氏名不詳. 続・組合せ範疇文法の漸進的構文解析, 2021. https://m-a-o.hatenablog.com/entry/2021/11/14/234710.

[13] Julia Hockenmaier and Mark Steedman. CCGbank: User's manual. 05 2005.

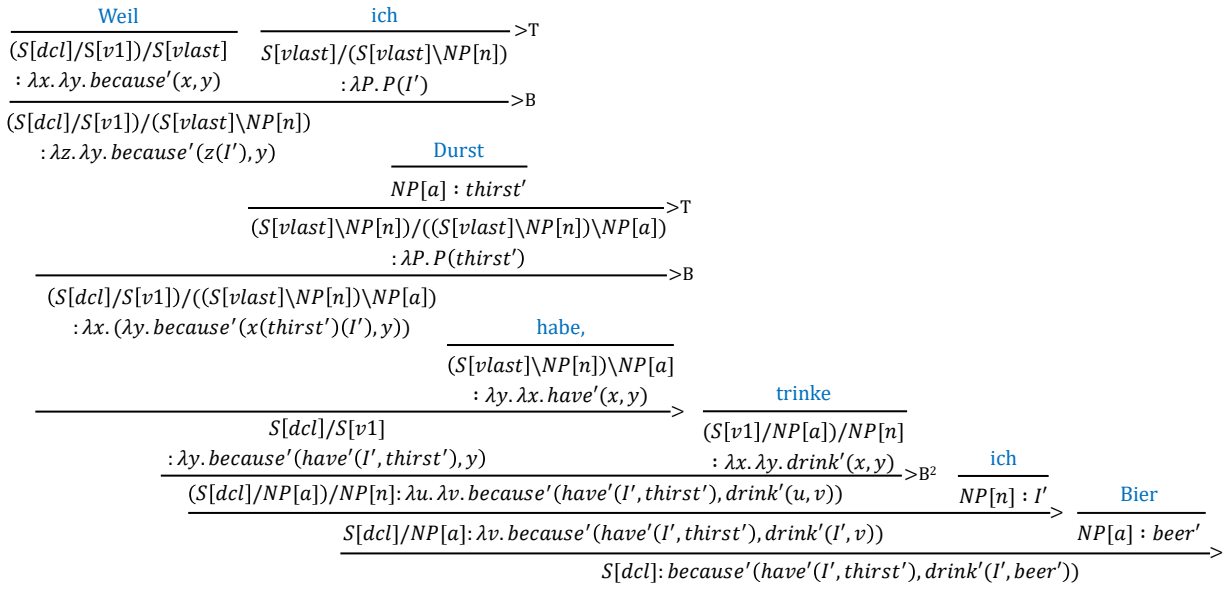

図 **8**　“Weil ich Durst habe, trinke ich Bier.” の漸進的導出

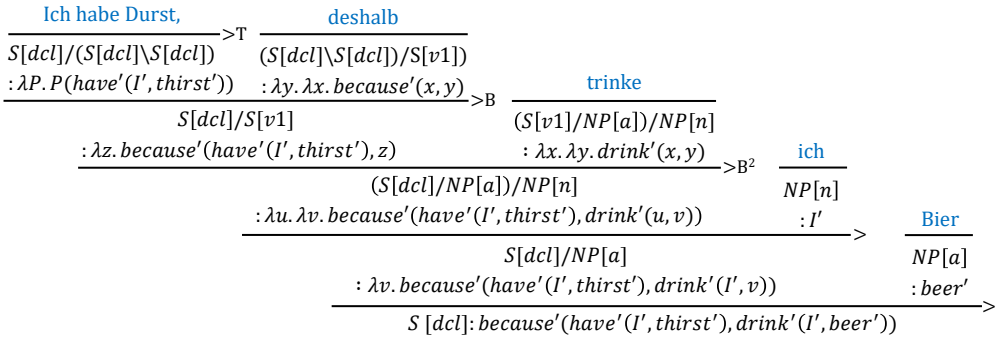

図 **9**　“Ich habe Durst, deshalb trinke ich Bier.” の漸進的導出．
先行文の “Ich habe Durst” は，“Ich trinke Bier” と同様の手順で漸進的に導出可能．

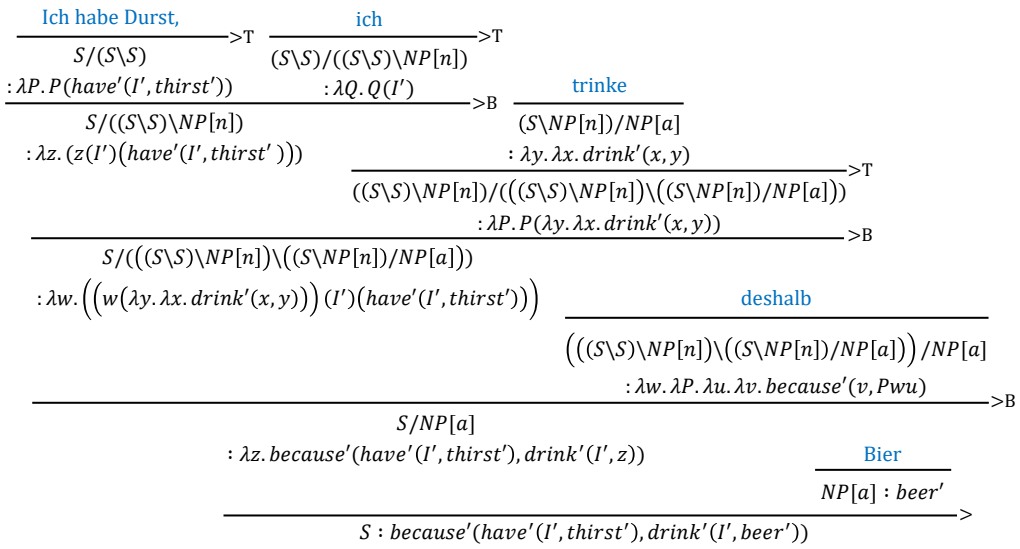

図 **10**　“Ich have Durst, ich trinke deshalb Bier.” の漸進的導出．S はすべて $S[dcl]$．