

自然言語インタフェースにおける対話的ユーザ語彙獲得機能

2F-2

谷 幹也 市山 俊治

NEC 関西 C&C 研究所

1 はじめに

自然言語インタフェースにおいて、エンドユーザが入力語彙を気にすることなく入力を行なうためには、入力文に含まれる未登録語情報を獲得するユーザ語彙獲得機能が重要である。

システムがエンドユーザへ適応する場合、1) 未登録語の獲得、2) 既登録語の言い換え、3) ユーザの言い回しの獲得の3つが存在すると考えられる。本稿では、未登録語の獲得に関して、前後の状況による推定の限界を明らかにするとともに、効率的に必要な情報を獲得する対話戦略を提案する。

エンドユーザとシステムの対話によって未登録語情報の獲得を行なう場合、システムのモジュール構成に関係なくエンドユーザの入力が冗長にならないようにする必要がある。獲得すべき情報、その選択枝あるいは誘導を行なうための提示可能な情報、情報の提示、選択、入力の方式の選択に関して一つの枠組を提案する。

また、現在筆者らが開発をすすめている自然言語インタフェース構築キット“IF-Kit”[1]上での実現方式を述べる。

2 自然言語処理の未登録語獲得

自然言語インタフェースでは、既知語から構成される文入力の解析、対象システムへのマッピングと同様に、未登録語を含む入力文に関してもその語彙を推定、対話による獲得を行なって統語構造、意味構造の作成、対象システムへのマッピングを行なう必要がある。未登録語に関する処理に着目した場合、未登録語の推定から登録までは図1のような流れとなる。形態素解析によって、未登録語区間の推定を行なう。未登録語情報の推定については、次章で詳しく述べるが、言語現象、解析構造、文脈情報、共起パターン、同義語辞書、デフォルト情報などを用いて行なう。未登録語の登録に必要な全ての情報を一意に推定できた場合は登録を行ない、一意に推定できなかった場合は、その絞り込んだ情報を用いてユーザへの対話を通じて未登録語情報の獲得を行ない、必要な情報を一意に決定できた段階で登録を行なう。

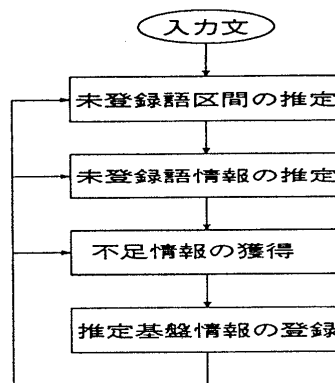


図1. 未登録語獲得のながれ

3 未登録語情報の推定

未登録語は、構文解析を行なうために、1) 体言系(名詞として扱えるもの)、2) 用言系(用言として扱えるもの)、3) その他に分類できる。それぞれの必須情報は、図2のように記述できる。必須情報の中の構文意味情報の推定は構文解析段階で次6つの情報を用いて推定を行なう。推定を行なった結果が一意に絞れない場合も候補情報として、未登録語情報獲得時に提示する。

	構文意味情報	対象システム情報
名詞系 未登録語	概念記号 意味分類	既知語への マッピング 単一概念 (同義語へのリンク)
用言系 未登録語	概念記号 活用ボタン 意味分類 必須格ボタン	複合概念 (マクロ概念の定義)
その他 未登録語	自立語への接続ボタン	なし

図2. 必須情報

言語現象 固有名詞として認定することが可能な接尾辞(氏、さん、市、研、部)や基本属性を表す特殊な名詞(名、長)などから推定が可能である。

解析構造 未登録語を含む解析構造と事例ベースとして蓄えた解析構造とのマッチングから未登録語の意味分類の推定を行なうことができる。また、等位構造や用言の格スロット条件から意味分類を荒く推定することができる。

文脈情報 前文までの話題や焦点のスタックから、文の構成要素の省略や指示詞の解釈を行なうが、その際にスタックされている構造と未登録語を含む構造のマッチングを行ない、そこから未登録語に関する意味分類を推定する。

共起ボタン 名詞連続や連体助詞による名詞句などの2つの名詞概念とその間の表層格、深層格の組合せを登録した共起ボタン事例ベースから、未登録語の意味分類候補を抽出する。

同義語辞書 未登録語の部分文字列から、同義語辞書を探索しヒットするものがあれば、その語の意味分類から推定を行なう。

デフォルト情報 用言の必須格スロット情報は、対象システム情報の獲得によって他の既登録語の用言と同義とみなされるまで”が格、を格、に格”をとるものとして処理を続ける。同じように用言の意味分類も”状態”や”動作”などのデフォルト情報として処理を続ける。

4 不足情報の獲得戦略

エンドユーザが言語的な専門知識や対象システムに対する知識を知ることなく未登録語の必須情報を設定するために、推定によって絞り込んだ情報を用いて次の3点を満たすように提示情報の作成と選択の誘導を行なう。

- 1) エンドユーザの最小の入力による必要な情報の獲得
- 2) システムとの自然な対話による情報の獲得
- 3) エンドユーザにわかる平易な形での情報提示

言語的な情報(活用ボタン、意味分類など)の獲得を行なう際には、例文によってそのいい回しが適切かどうかを判断してもらうなど専門知識を使用しない誘導を行なう。次の4つの機能を用いることにより、このユーザへの誘導の提示と問い合わせを行なうことができる。1) 確認(情報の提示とyes/noの回答), 2) 一意選択(複数の選択枝から一意に選択), 3) 複数選択(複数の選択枝から一意に選択), 4) 自由入力。

前章で述べた必須情報のうち、一意に決定できた情報とそれ以外を各モジュール間で伝達する。特に対象システム情報によって既知の語彙へのマッピングが決まれば構文意味情報の意味分類や格スロットボタンなども決定できるので、対象システム情報の獲得後、再度ユーザ辞書への反映を行なう。

5 自然言語インタフェース IF-Kit での実現

現在筆者らが開発している自然言語インタフェース構築キット”IF-Kit”[1]では、自然言語インタフェースを1) 構文解析、2) 文脈解析、3) タスク解析、4) 対象言語生成部の4つのモジュールに分割している。

構文解析 形態素解析部で辞書引きできなかった区間を未登録語としてとらえ、この未登録語を、構文解析部は解析

ルールを用いて用言系、体言系、その他に分類し、それぞれの場合について図3のような処理を行なっている。

文脈解析 文脈解析では、前文の話題構造と現在の構造とのマッチングをとることで未登録語の意味分類の候補を抽出することができる。

言語現象	推定	獲得
体言系		
名詞連続構造 名詞句構造 助詞接続	未登録語でない名詞からの推定 共起ボタン事例からの推定 体言系の推定	タスク解析でのマッピング 獲得時に意味分類の獲得
等位構造 用言の 格スロット条件	意味分類の宛い決定	タスク解析でのマッピング 獲得時に意味分類の獲得
用言系		
活用語尾接続 助動詞接続	助動詞,活用語尾表層 からの活用ボタンの推定 意味分類,格スロットの デフォルト推定	タスク解析でのマッピング 獲得時に格スロットの獲得 例文提示による活用ボタンの獲得
その他	字種,長さ,自立語からの 接続ボタン推定	例文提示による接続ボタンの獲得

図3. 構文解析による推定

タスク解析 タスク解析部では、入力された構文意味情報を対象システム内の語彙にマッピングしている。未登録語に関しては、ユーザ知識DBへ登録し、既存のネットワーク形の対象領域知識ベース上の概念へのリンクをユーザ知識DBに登録することで登録を行なう。この時ユーザからの入力の最小とするために、次のような入力の誘導を行なう。最初に、[1]無視する、2) データとする、3) 言い替える]の3つから選択してもらい、2) に対しては、その語をデータとして持つようなフィールドをネットワーク内で絞り込み提示する。3) の「言い換える」では自由入力によって入れ直してもらおう。一度未登録語として獲得されたものは、既知語概念か既知語概念のネットワークに対して同義語リンクで繋ぎ、同じ語に関しては同義語を処理するように言い替えの確認を行なう。言い替えの確認は文脈情報がクリアされるまでストックされ一度確認されたものは2度と確認されることはない。

6 おわりに

自然言語インタフェースで重要となる未登録語獲得をエンドユーザへの負荷最小という立場から行なう方式について提案し、IF-Kit に実装を行なった。現在、エンドユーザへの提示情報の推定、選択を行なう方式に対して、評価を行なっている。今後は、インタフェースとのレスポンス時間など他の方面からのエンドユーザ負荷の軽減方式について考察する予定である。

【参考文献】

- [1] 谷幹也, 飯野香, 山口智治, 市山俊治: 自然言語インタフェース構築キット:IF-Kit, 信学技法 NLC91-62, 1991.