

日本語文解析器および生成器のサーバ化に関する研究

5B-2

原田 雅之, 唐澤 博

山梨大学 工学部 電子情報工学科

E-Mail: (harada,karasawa)@opal.esi.yamanashi.ac.jp

1はじめに

自然言語を入出力インターフェースとするシステムを多数のユーザが使用する環境では、ユーザ数分の日本語文解析器および生成器を逐一動して用いるのはメモリ効率及び実行効率が悪い。そこで、マルチユーザ対応の漢字変換システムWnnの場合と同様な考え方で、解析器・生成器をサーバ化して、各ユーザ・アプリケーションをクライアントとする構成のシステムを開発した。クライアントは文解析サーバから、ユーザの入力した日本語文が解析されて意味構造になったものを受けとる。一方、F構造を文生成サーバに送ることによって日本語出力文を受けとる。双方の過程で、文脈情報も受け渡しが行なわれて更新される。この方式を自然言語対話Agentシステムに適用した結果を報告する。

2 日本語文解析器および生成器

2.1 日本語文解析部

ユーザからの自然言語での入力をシステムに理解できる構造に変換するモジュールである。まず、入力日本語文はLAX流の形態素解析部において形態素列に変換される。次に、構文・意味解析、文脈解析部[1][2]においてLFG(語彙機能文法)に変換される。さらに、意味構造生成部でUSF(発話意味構造)に変換され、意図解析部においてUGF(発話意図構造)に変換される。最後に、社会関係解析部[3]においてシステムの行動が決定されて適切な実行スキーマが呼び出される。

2.2 日本語文生成部

システムの発話内容を自然言語に変換するモジュールである。実行スキーマからUGFを受け取り、意味生成部においてUSFに変換される。次に、構文生成部においてF構造(語彙機能構造)に変換され、形態素生成、代用・省略処理などの文生成部を経て日本語文が出力される。

3 サーバ・クライアントモデルの構成

自然言語対話Agentシステムは図1に示す構成となる。まず、トップ・スキーマが起動され、ユーザに対して開始・終了の挨拶をするものとなっている。

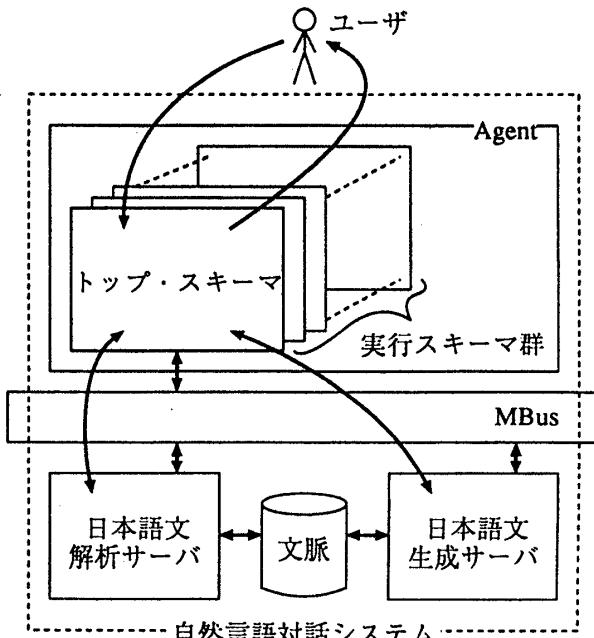


図1: 自然言語対話 Agent システム

3.1 サーバ化の目的

逐一動型の構成の場合、ユーザ数分の非常に巨大な日本語文解析器および生成器を毎回立ち上げて用いていることから、メモリ効率及び実行効率が悪いことが多い。そこで、解析器・生成器をサーバ化して、各ユーザ・アプリケーションをクライアントとする構成のシステムの場合では、メモリ効率及び実行効率が良くなると考えられる。さらに、1つのサイトにおいて数多くの言語が扱われていることは稀であることから、そのサイトで扱われている言語分の文解析・生成サーバ対のみが常駐していれば良いことになる。

3.2 日本語文解析部の実現方法

3.2.1 クライアント

ユーザの入力する文章を標準入力より読み込み、文脈情報と共に解析器専用入力スプールに置いて、ユーザが扱う言語の文解析サーバに対して要求を出す。そのサーバから終了信号が送られて来たら、解析器専用出力スプールに置かれた更新後文脈情報と発話意図構造を受け取る。

A Server-Client System for Japanese Text Analysis and Generation

Masayuki Harada, Hiroshi Karasawa

Yamanashi University

4-3-11 Takeda, Kofu, Yamanashi 400, Japan

3.2.2 サーバ

文解析クライアントから要求を受けると、解析器専用入力スプールに置かれた入力文と文脈情報を取り込み解析する。解析し終ればその結果である発話意図構造と更新後文脈情報を解析器専用出力スプールに置き、クライアントに対して終了信号を送る。なお、解析中での他クライアントからの要求は解析器専用入力スプールに溜まっていることから、現在の解析が終了次第取り掛かることになる。

3.3 日本語文生成部の実現方法

図2に日本語文生成部の動作図を、図3に動作例（この場合はF構造から日本語文の生成）を示す。

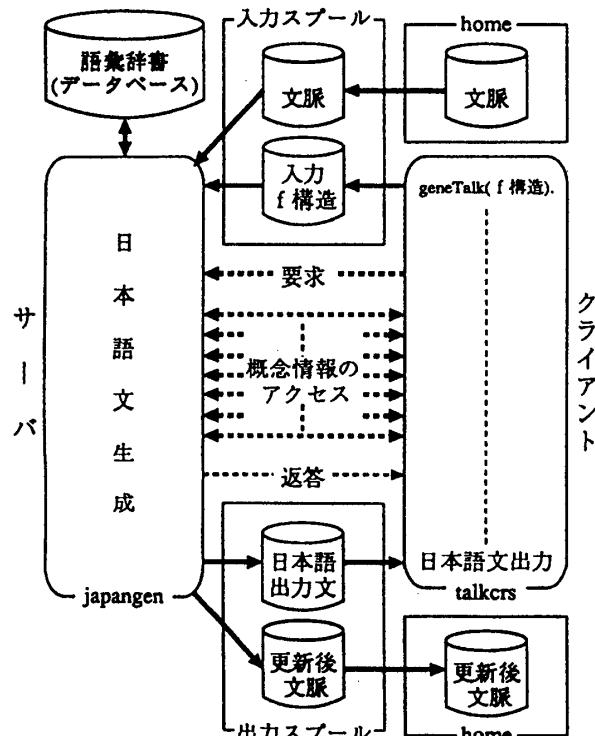


図2: 日本語文生成部の動作図

3.3.1 クライアント

実行スキーマが出力した発話意図構造を、文脈情報と共に生成器専用入力スプールに置いて、ユーザが扱う言語の文生成サーバに対して要求を出す。そのサーバから終了信号が送られて来たら、生成器専用出力スプールに置かれた更新後文脈情報と出力文を受け取り、その出力文を標準出力へ書き出す。

3.3.2 サーバ

文生成クライアントから要求を受けると、生成器専用入力スプールに置かれた発話意図構造と文脈情報を取り込み生成する。生成し終ればその結果である生成文と更新後文脈情報を生成器専用出力スプールに置き、クライアントに対して終了信号を送る。なお、生成中での他ク

ライアントからの要求は生成器専用入力スプールに溜まっていることから、現在の生成が終了次第取り掛かることになる。

3.4 MBusによるプロセス間通信

クライアント～サーバ間通信はメッセージ・プラットホーム MBus[4] を経由する。MBusとは、イリノイ大学で開発されたプロセス間のメッセージ交換を効率的に媒介するソフトウェア・バスである。

```
nemacs: emacs @ lapis
?->genTalk([interJ=[lexID=1415,
inst=0],
subj=[lexID=1726,
inst=2,
theme=yes,
pers=1,
num=sing,
territory=そ],
objgo=[lexID=1720,
inst=1,
theme=no,
pers=2,
num=sing,
territory=こ],
obj=[lexID=4507,
inst=10,
theme=no,
pers=1,
num=sing,
territory=そ],
lexID=5947,
inst=65,
mood=[接与,対者丁寧,意志],
territory=そ]).
はい、本を読んであげましょっ。
?->
```

図3: 日本語文生成部の動作例

4 おわりに

正確な動作確認の為、連続要求実験を行なったが正確な返答を全て返している。ただ、サーバの能力上待行列が出来てしまう。その待行列を最小限に抑えることが今後の課題である。この対策法としては、サーバ側のアルゴリズムの高速化などが考えられる。

参考文献

- [1] 小野博隆, 唐澤博: 補完候補の優先順位の決定, 情報処理学会第48回全国大会講演論文集, 4R-3, pp.181-0.582(1994.3).
- [2] 桜井啓司, 唐澤博: コネクショニズムを適用した文脈理解, 情報処理学会第50回全国大会講演論文集(3), 6R-9, pp.141-0.542(1995.3).
- [3] 林亮, 唐澤博: 社会関係解析による聞き手行動プランの生成と心理モデルの研究, 情報処理学会第46回全国大会講演論文集, 8B-8, pp.205-206(1993.3).
- [4] 武田英明, 飯野健二, 西田豊明: 知識コミュニケーションにおける知識共有メカニズム, 人工知能学会第8回全国大会論文集, pp.279-282, (1994.6).