

SE 作業工程における設計用語の入力支援

鄭育昌*, 長瀬友樹*, 笈田佳彰†, 岡田伊策†

株式会社富士通研究所 メディア処理研究所*

富士通株式会社 S I 技術本部 システム技術統括部†

{cheng.yuchang, nagase.tomoki, oida.yoshiaki, isaac-okada} @jp.fujitsu.com

1 はじめに

情報システムの企画、開発、運用・保守は、顧客、SE（システムエンジニア）を含めた多くのプレイヤーによる共同作業である。プレイヤー間で作業を進め、その過程の品質を確保するために、共通認識を合わせるための共通の拠りどころが必要である。例えば、SDEM®¹のような企画から運用・保守にいたる各工程に必要な作業項目が記述された作業工程基準に従うことで、作業の漏れを防ぎ、かつ効率的にプロジェクト運営を実現することができる([1])。

しかし、作業工程におけるコミュニケーションの多くは文書ベースのため、文面の曖昧性がある場合、設計書のレビュー工数の増加、またはプログラムの実装漏れなどが発生することがあり、プロジェクトの円滑な運営に支障が出る。特にシステムの機能名や項目名など、開発工程においてあいまい性が許容できない設計用語²に対し、文書における表記の統一が求められている。

本研究では、情報システムの開発における意思疎通の障害の要因である設計用語の不統一を解消するため、SDEM®に準拠した SE 作業工程の流れを調査し、各工程の用語に関する課題、及び問題解決の技術要件について整理・考察した。

前記の考察結果に基づき、設計用語の不統一が起きやすい設計書の作成工程にリアルタイムかつインタラクティブに設計用語の誤入力を指摘できる機構を開発した。本稿では、提案の設計書入力支援機構の構成と使用イメージを説明する。

この入力支援機構を、現場で頻出するプロジェクトの設計用語（機能名、項目名）の表記不統一の事例に適用して評価した結果、指摘正解率 86%を達成した。

本機構の利用により、文書レビューの工数削減またはプログラム実装時の仕様誤解や実装漏れの削減

などの効果が期待される。今後は、開発した設計用語支援機構を SE 現場に提供し、SE 作業工程において本機構の利用効果を評価する予定である。

2 SE 作業工程分析と用語関連の課題抽出

本章では、当社の SE にヒアリングを行うことで整理した SE 作業工程の流れと用語関連課題について説明する。図 1 は、SDEM®に準拠した SE 業務における各工程のプレイヤー、作業項目、用語の状態、用語の方針および用語の問題/課題を整理したものである。なお、図 1 の各工程の略称（VP~OT）および工程名は本稿 APPENDIX の「表 3: SDEM 工程名と略称の対応表」に対応する。

図 1 は、時系列としては図の左から右へ進んでいる。熟練者 SE の知見により、SE 作業工程を用語の定着（設計用語表現の不統一が発生しやすさ）状態を基準に、以下は「用語収束期」、「用語定着期」、および「用語安定期」の三段階を区別して各段階の状況と課題、および課題解決に必須な技術を説明する。

2.1 用語収束期

概要：SE 作業の初期工程の段階（VP-RD）に相当する。上流 SE が顧客の要件を聞き出すため、顧客側（経営者と情報システム部門担当者など）との会話を行い、プロジェクトに関するシステム要件定義のドキュメントを作成する。

設計用語の状態：顧客や業界の用語、自社（SE 会社）の用語、既存システムの用語などが並立して未統一の段階であり、用語系の作業方針は用語のすり合わせ及び今後の作業のための標準用語の策定である。

課題：既存システム用語、他業界の用語が多種多様のため、業界関係者以外は直感的に用語を理解することができない場合がある。そのため、用語の収束・統一および意識・概念合わせの支援は用語系の課題である。

想定技術：未統一の標準用語の収束に支援するため、各業界（会社）が使う用語の特有意味を記述する言語リソース（辞書や用語集）、用語一般的な意

¹ SDEM®は、富士通の企画、開発、運用・保守、品質保証活動の基本的な考え方を示した標準プロセス体系であり、すべてのステークホルダーの共通認識を形成するための地図である。

² 本稿では「設計用語」が情報システムのデータ項目名や機能名を指す。

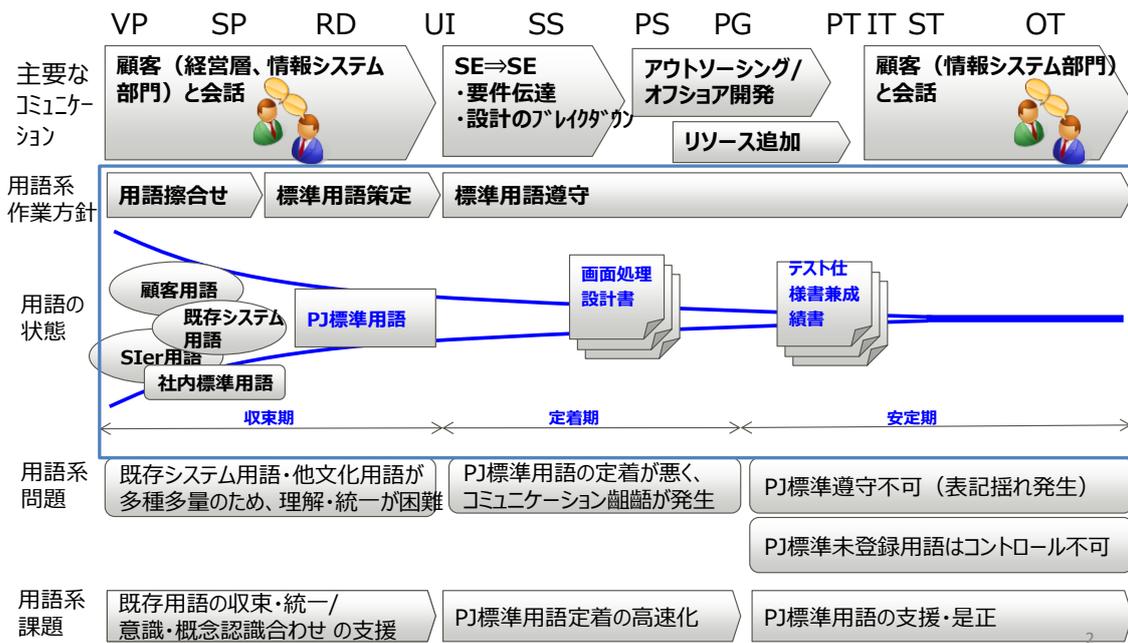


図 1: SDEM®に準拠した SE の作業工程の流れと用語関連課題の整理

味と区別した特殊専門用語の同義・類義語の識別技術が必要である。

2.2 用語定着期

概要：SE 作業の開発工程（UI-SS, PS-PG）に相当する。開発工程に入り、システム要件に基づくシステム設計などを行い、設計書を大量に生成する。

設計用語の状態：初期工程で顧客の要件を聞き出したことと共に、プロジェクト用語の収束を行なって、プロジェクト標準用語の定義も行った。開発工程に入る本段階では、用語系の作業方針は策定した標準用語を確実に遵守することである。

課題：設計書の作成などの場合、プロジェクトの標準用語の定着が悪くなると、コミュニケーション齟齬が発生し、レビュー工数の膨大化や実装漏れなどの障害が出る。そのため、外注先も含めて、プロジェクト関係者のプロジェクト標準用語の定着化が効率よく達成することが課題である。

想定技術：本段階では、新たに策定されて不慣れた設計用語を使って、SE やプログラマーが設計文書を大量に作成する。以降の工程に悪影響を及ぼさないため、設計用語の不統一が発生する場合、リアルタイムで検知、用語を修正するための支援技術が必要である。

2.3 用語安定期

概要：プロジェクト収束の工程（PS-PG, PT-OT）に相当する。関係者（SE、プログラマー、外注先など）が標準用語を元に設計ドキュメントやマニュアル

ルなどの作成・修正・追記作業に入り、納品に向け、再び顧客側（情報システム部門）とのコミュニケーションを行う。

設計用語の状態：標準用語が定着し、新規文書の作成が比較的になくなる。

課題：熟練者が正確に設計用語を使用するが、新規参加者がドキュメントを修正する際、未登録が使用される状況が発生する。プロジェクト標準用語の校正支援が課題である。

想定技術：本段階は修正された既存文書に対し、入力ミスなどの表記ゆれや未登録語をバックグラウンドで一括検出・修正する技術で対応可能である。

2.4 考察

用語観点からの SE 作業の分析により、各工程での用語関連の課題が明らかになった。上流工程の「用語収束期」では、顧客の要件定義は収束しておらず、複数の用語リストが並立している。したがって、このフェーズの表現・用語統一の作業は、顧客との会話でプロジェクト標準用語を定義することが中心である。下流の工程である「用語安定期」では、設計用語の正しい使用が定着していることから、表現・用語の統一は、それほど大きな問題とは認識されていない。この工程では、作成された文書の単位で一括して表記ゆれをチェックできる、従来型の文書校正技術が有効であると考えられる。

一方、「用語定着期」は情報システム実際の開発作業に相当し、情報システム開発用のフレームワー

表 1: 本稿の調査対象の SE ドキュメント（定着期以降）の概要

	合計
プロジェクト数	5
ドキュメント数	9,169
文字列数（行数）	3,929,072
名詞（複合名詞込み）抽出数	135,795
同義表現ありの名詞数	18,916
同義表現グループ数	5,834

表 2: SE ドキュメントにおける同義表現グループの事例

No.	事例
1	顧客名,お客様名,顧客名称,お客様名称
2	ドキュメント,文書,書面,記録
3	カラー名,色名,カラー名称,色名称
4	店マスタ,店別マスタ,店舗マスタ,店舗別マスタ
5	日別店別商品分類別予算テーブル,日別店舗別商品分類予算テーブル
6	粗利金額合計,粗利合計金額,累計粗利金額,粗利金額集計
7	売上累計数量,売上数量合計
8	売上原価合計金額,売上原価金額合計,累計売上原価金額,売上累計原価金額

く³上で項目名と機能名を含む設計書や仕様書を多数作成するフェーズである。ここでは、統一された設計用語の使用が定着しておらず、項目名などの用語の不統一が発生し、これが後段のシステム障害を引き起こす原因になる場合がある。そのため「用語定着期」では、文書の作成時に用語の誤用を即座に指摘してくれるようなシステムが有用であると考えられる。

「用語定着期」における設計用語の誤用について調査するため、我々は5つの情報システム開発プロジェクトの SE ドキュメントを入手して用語の同義性について調べた⁴（表 1）。表 1を見ると、約9千件のドキュメント（4百万行の文字列）に、複合名詞を含む約13万の名詞および複合語が含まれることがわかる。これらの名詞および複合語の構成単語の意味を調べた結果、計5834個の同義グループ（計1万8千個の名詞）を抽出し、SEドキュメント

³ 例えば、ソフトウェアの設計、開発、運用、保守を支援するアプリケーションフレームワーク製品

「INTARFRM®」を用いた開発工程の場合
<http://www.fujitsu.com/jp/solutions/infrastructure/dynamic-infrastructure/afw/>

⁴ 名詞の同義表現の調査手法は、3.1節の学習フェーズの処理と同一である。

に使われる名詞の約13%が同義表現を持つという結果になった。

表 2 は、抽出された同義表現の事例を示している。例えば、表 2 の No.3 のように、プロジェクト A の「カラー名」をプロジェクト B では「色名称」と表現する場合、プロジェクト B の SE がプロジェクト A に参入する場合、本来「カラー名」とすべきところに「色名称」という表記を使ってしまう可能性がある。

このように、人材の流動性に対応した用語統一支援は、「用語定着期」における SE 作業の大きな課題の一つである。昨今ビジネスシステムのトレードは SoR⁵から SoE⁶に移行しつつあり、業界横断、異業種連携参入の企業が増加するなかで、SE が過去の担当と異なる業界の案件に携わるケースが増えている。プロジェクト毎の標準用語に慣れるまでには長い時間を要するため、業界に不慣れな SE に対してドキュメント作成時の同義語をチェックし標準用語へ導くことができるシステムに対するニーズが大きい。

本研究の設計用語の表現統一技術の適用先は、SE 作業工程「用語定着期」を想定する。SE が設計書を作成する時、リアルタイムで表現不統一の可能性のある複合名詞を検出し、業種で標準的に使われている単語への修正をレコメンドする。これによって設計書の不統一表現の発生を防ぎ、情報システムの開発作業工程の効率を向上させることが可能であり、「SE の支援力・即応力の向上」にも寄与すると考える。

3 設計用語の入力支援

我々は、前章の分析調査で得た知見を元に、リアルタイムでインタラクティブな設計用語の表現統一の入力支援機構を提案した。本章ではその機構の機能、構成、及び利用イメージを説明する。

提案の設計用語入力支援機構は、まず、学習フェーズで複数のプロジェクトから不統一表現の可能性のある同義表現の収集と学習を行い、同義表現データベースを作成する。利用フェーズではユーザの入力と同時にリアルタイムで不統一表現の可能性のある箇所をレコメンドすることで不統一表現の発生を文書の作成段階で防止する。

本研究では、以下事例のように、単一の概念記号 ([2]) を持つ「単語」と複数の概念記号を持つ「複合名詞」を区別する。概念記号とは、ルールベース機械翻訳における単語表記の意味を表すものである。不統一表現であるかどうかの判断が困難な「単語」

⁵ SoR (System of Record): 企業内のデータを記録し、業務処理を行う従来型の業務システム

⁶ SoE (System of Engagement): 人やモノを含めすべてがつながる新たなシステム

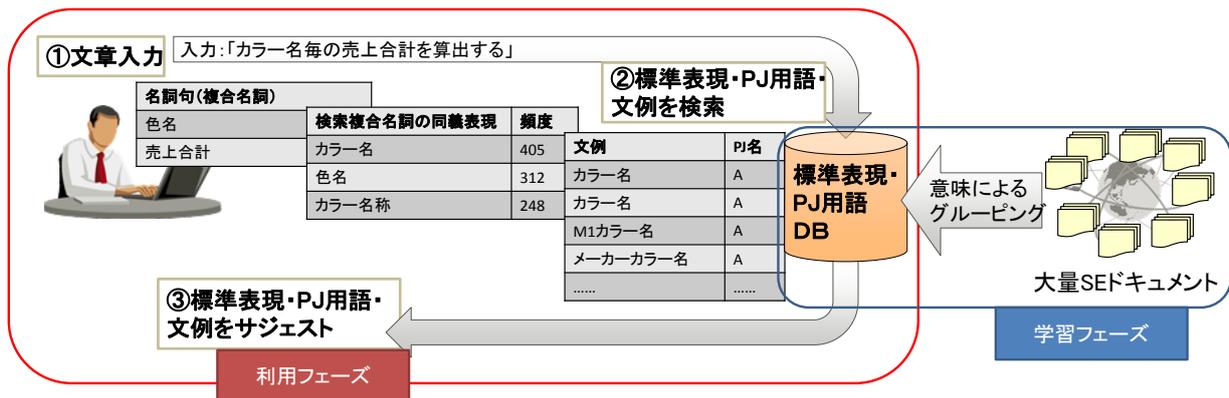


図 2 : SE 作業工程における設計用語入力支援機構の利用イメージ

を処理対象から除外し、以降の不統一表現の処理対象は「複合名詞」のみに限定する。

- 単語例：システム、system
- 複合名詞例：情報システム、インフォメーションシステム

3.1 学習フェーズの処理

学習フェーズの処理は、大量 SE の設計書から、不統一表現を収集することである。図 3 は設計用語入力支援機構の学習フェーズの構成を示す。SE が作成した大量設計書ファイルから、文字列を抽出し、各単語の概念記号の意味関係を示す概念構造 ([2]) の解析を行い、概念構造から複合名詞を抽出する。抽出された複合名詞に対し、複合名詞が持つ概念

記号を比較し、同じ概念記号を持つ複合名詞を同一クラスターにまとめる(概念記号の順番を考慮しない)。概念記号は表記の意味を表すものであり、同じ意味の表記の概念記号が同じである。例えば、表記「名前」、「名称」、「ネーム」、「名」の概念記号は同じ[name]である。複合名詞の事例は以下の「売上金額総合計」と「累計売上金額」のように、複合名詞の構成概念記号が同じのものは同一のまとまりにグルーピングする。

- 複合名詞同義例 1 :
色名 = [color]+[name]
カラー名称 = [color]+[name]
- 複合名詞同義例 2 :
売上金額総合計 =
[PROCEEDS]+[AMOUNT OF MONEY] + [TOTAL]
累計売上金額 =
[TOTAL]+[PROCEEDS]+[AMOUNT OF MONEY]

以上の処理を経て、複数プロジェクトから収集した大量設計書の同義語表現データベースが構築される。

3.2 利用フェーズの処理

設計用語の表現不統一の発生を防ぐため、文書の作成段階で入力サジェストの方式でインタラクティブ文章作成を支援することが望ましい。前節の学習フェーズの処理を経て、不統一表現の可能性がある同義語表現データベースが作成される。利用フェーズでは、設計書の作成者が不統一表現の可能性がある表現を入力する際に同義表現の存在を作成者に提示し、不統一表現であるかどうかの確認を促す。この目的を達成する利用方法 (UI) として、本研究は IME と連携して、入力文字列の同義表現を指摘するインタラクティブな設計用語入力支援の方式を提案する。

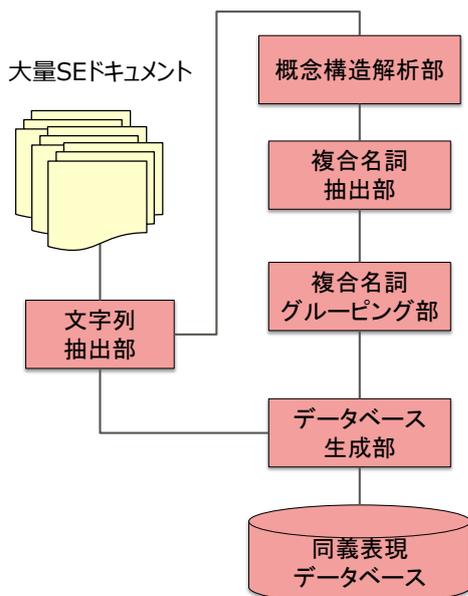


図 3: 設計用語入力支援機構の構成 (学習フェーズ)

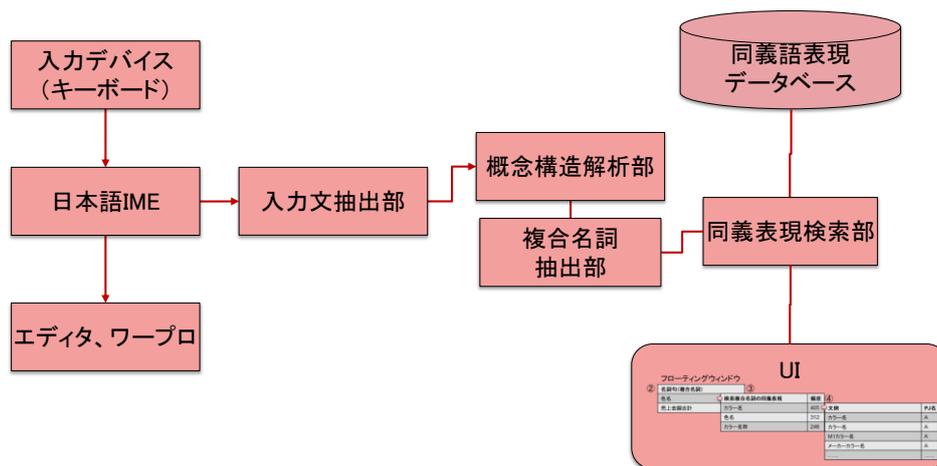


図 4:設計用語入力支援機構の構成 (利用フェーズ)

図 4 では、利用フェーズの処理を示す。本入力支援機構は IME と連携し、文書の作成者 (SE) が入力デバイスで文字列を入力する際、入力文抽出部で入力文字列を抽出する。その後、入力文字列が入力支援機構のデータベース検索部に渡され、入力文字列がデータベースに登録された不統一表現である場合、不統一表現の箇所および同義表現を文書の作成者に提示し、利用者に入力内容の再考を促す。

図 5 は上記の処理を実現する UI のイメージである。ワープロソフトで文字列を入力する際、以下の処理を行なって、同義表現を提示する。

- ①ユーザ入力から解析対象文の抽出
- ②①の入力文から複合名詞を抽出、フローティングウィンドウに表示
- ③ユーザが②から選択した複合名詞と同義である既存表現を検索、結果を新しいフローティングウィンドウに表示
- ④ユーザが③から選択した同義表現の文例を新しいフローティングウィンドウに表示

例えば、ユーザが「表 3 に色名と売上金額を登録する」を入力する際、まず、入力文から複合名詞「色名」と「売上金額」が検出され、②のフローティングウィンドウに表示される。

ユーザは複合名詞「色名」に対して判断を行うため、マウスを②の「色名」に移動し、「色名」の同義表現が③のフローティングウィンドウに表示される。この事例では、「色名」の同義表現に複合名詞「カラー名」と「カラー名称」が過去の設計書に使われたため、「色名」が同義表現の不統一表現の可能性がある。ユーザに同義表現を表示することで、複合名詞「色名」の利用可否を再考するようにユーザに促す。

更にユーザが各複合名詞の出現箇所と文脈を参照するため、マウスを②の複合名詞合わせると該当同義表現が過去の設計書に出現した文例がフローティングウィンドウ④に表示される。この UI で表示された情報を元に、入力された不統一表現の可能性がある複合名詞に対して訂正することができ、不統一表現の出現が文字列の入力段階で予防できる。

また、本 UI の利用形式 (入力時のリアルタイム指摘) により、ユーザ (設計書作成者) が最初に不統一表現が入力された時にその場で指摘されたため、以降の誤用を自然に回避する効果がある。

3.3 設計用語入力支援機構の主観評価

提案の設計用語入力支援機構の適用効果は、設計用語の不統一の発生を防止することによるレビュー工数削減などの効果が期待される。本研究では、まず、提案機構が設計用語の不統一事例に対する検出精度を計測した。

検出精度は、表 1 のプロジェクト開発工程に携わった熟練者 SE の視点で、指摘すべき同義表現に対する本機構の推奨の適切さを判定することで評価した。設計用語の不統一は顧客の要望により同義表現を使い分ける場合もあるため、推奨の適切さは熟練者 SE の主観で判定する必要がある。

熟練者 SE が表 1 の SE ドキュメントを元に、開発工程に出現しやすい設計用語の不統一表現を入力し、本機構が入力に対して適切な設計用語を推奨できるかどうかを確認する。例えば、

図 5 の事例を入力し、仕様通りに「色名」と「売上金額」が検出、各自に本プロジェクトでは同義表現の設計用語が存在することを示せば正解と判断する。

その評価結果は、想定した 49 件の事例の内、適切な推奨を提示できたのは 43 件であった

①エディタ、ワープロの入力域

表 3 に色名と売上金額を登録する。

② 名詞句 (複合名詞)	③ 検索複合名詞の同義表現	④ 頻度	④ 文例	PJ名
色名	カラー名	405	カラー名	A
売上金額	色名	312	カラー名	A
	カラー名称	248	カラー名	A
			M1カラー名	A
			メーカーカラー名	A
		

図 5: インタラクティブな設計用語入力支援の UI 構成

(指摘率: 88%)。指摘できなかった事例 (6 件) は、形態素解析などの失敗で、複合名詞を正確に抽出できなかったことが原因である。

4 まとめ

本研究では、情報システムの開発における意思疎通の障害である設計用語の不統一を解消するため、SE 作業工程の現状を調査し、各工程の用語に関する課題調査、及び、問題解決の技術要件について整理・考察した。前記の考察結果に基づき、設計用語の不統一が起きやすい設計書の作成工程にリアルタイムかつインタラクティブに設計用語の誤入力を指摘できるツールを開発した。このツールを、現場で頻出するプロジェクトの設計用語 (機能名、項目名) の表記不統一の例に適用して評価した結果、指摘正解率 86% を達成した。

今後は、提案システムを開発し、現場の SE のトライアルにより、本技術の適用による不統一表現の削減効果を定量的に評価する予定である。

APPENDIX

表 3: SDEM® 工程名と略称の対応表

工程名	略称
情報化構想立案	VP
システム化計画	SP
システム要件定義	RD
ユーザーインタフェース設計	UI
システム構造設計	SS
プログラム構造設計	PS
プログラミング	PG
プログラムテスト	PT
結合テスト	IT
システムテスト	ST
運用テスト・移行	OT
運用・保守	OM

参考文献

- [1] 室中健司、原直朗, 2012. システム構築の標準プロセス体系: SDEM (特集 新たな時代の SI 技術). *Fujitsu* 63, 2 (2012/03), 193-199.
- [2] 大倉清司、潮田明, 2014. 複雑な文に対応した意味構造検索システムの開発. In *言語処理学会第 20 回年次大会*, 北海道札幌市.