

インターネット上の求職票に見られる業務記述の分析

—SOHO ワーカは受発注のために何を重要と考えているのか—

大熊智子 上野裕一

富士ゼロックス（株） 中央研究所

Analysis of Work Descriptions in Job Advertisement on the Internet

—What is regarded as important for SOHO workers to get the employment opportunity?—

Tomoko Ohkuma

Yuuichi Ueno

{Ohkuma.Tomoko, Yuichi.Ueno}@fujixerox.co.jp

Fuji Xerox Co., Ltd. Corporate Research Center

The purpose of this paper is to investigate :1)significant word set to be required for making contract on business, 2)features involved in the set, and 3)difference of the set among the types of business.

We studied SOHO workers' work descriptions through applying morpheme analysis and calculation of frequency for extracting word set from about 4600 descriptions on the Internet. As the result, lexicon of ten types of business were extracted.

The above words were observed to examine what features were involved in the work descriptions. Through the observation, we found that the descriptions can be characterized with the items: "Operation", "Object", "Domain", "Purpose", "Feature", "Attitude", "Skill", "Experience", "Requirement", "Price", "Delivery date", and "Tool and Means".

Moreover, assuming the attribute be an independent variable, and the type of business be a dependent variable, we converted work descriptions into binary data. Then, we applied category regression analysis to find the significant factors in three types of business "Translation", "Accounting", and "Software development and SE".

1 はじめに

1.1 背景

我々は、インターネットのような開放型ネットワーク上での業務連携を支援するためのシステム構築について、研究を進めている。[mas00] では、業務提供者間での業務連携を自動的に行なうことを可能にする業務記述のモデルを提案した。これは、[Mar97] における活動分類にもどづく 8 つの基本動詞を操作名とし、その操作対象となるあらゆる名詞を操作対象名とするものであり、受発注関係はタスク・サブタスク関係として表現される。

開放型ネットワーク上での業務連携サービスの利用者の一例として、我々は SOHO ワーカを想定している。SOHO ワーカは業務が発生した場合、ネットワーク上に分散した環境で活動を行なうため、このような業務連携機能は非常に有効であろうと思われるからである。そこで、本研究では SOHO ワーカが SOHO ワーカ登録サイトに掲示している求職票に対して分析を行なう。このような求職票は、面識のない発注者にむけ

て、自分の従事する業務内容や特徴が必要最小限の文章で記述されている。つまり、求職票は、ワーカの活動のエッセンスであると言っても過言ではない。この求職票を分析すれば、SOHO ワーカが業務活動において実際にどのような語をどのように使用しているのかを推測することができると思われる。

1.2 目的

本研究では上述の二つの問題、すなわち適切な業務表現を行なうための語彙の定義と記述項目について提案する。また、記述表現について各業種ごとの特徴についても考察する。

各業務において、従事者が受発注関係を成立させるのに必要十分な語彙はどのような語集合によって定義されるのか、またそれらはどのような属性を持つのか。そして、その性質は業務によってどのような相違があるのか、あるいはすべての業務は同質の記述スキーマを持つのか。これらを明らかにすることが、本研究の目的である。

この論文は以下のような構成になっている。2 章では複数の SOHO 関連のサイトのうち、最も登録者数の多かった SOHOO! Japan¹と Creators-gate²に掲載された求職票から半自動的に語彙の抽出を行ない、その特徴について考察する。3 章では前章で抽出した語彙を分類し、求職票の記述項目を抽出する。さらに、4 章では、業種間における求職票の構成の差異について、数量化 II 類による分析を試みる。これによって、業種によってどの項目が重要な項目となっているのかを明らかにする。

2 語彙の抽出

2.1 手続き

SOHO ワーカーの担う業種は、実に多岐に及ぶ。そこで、SOHO 業務支援サイトの業務ディレクトリを調査し、SOHO ワーカーの従事者が比較的多かった、「翻訳」、「ライター」、「ワープロ入力」、「ソフト開発・SE」、「Web 製作」、「DTP 関連」、「デザイン」、「CAD 関連」、「経理・事務」を対象とする。

求職票からの重要語の抽出は、作業の前段階の作業を自動的に行い、後段階の作業を人手で行った。まず、対象とする求職票に対し chasen ver2.02[mat99] を用いて形態素解析をかけ、動詞と名詞のみを取り出す。次に、情報検索における単語の適合性評価基準の一つである tfidf の値を計算し各業務ごとに閾値をヒューリスティックに設定し、重要語を抽出する。ここまでを自動的に行ない、さらに下記のような作業により、語彙の精度を向上させた。

形態素解析に失敗した語の修正／排除 形態素解析の段階において、語の切り出しに失敗した語が含まれている場合がある。例えば「検定」という語が「検」「定」という 2 語に分離している場合などである。しかし、たとえ間違った語でも、tfidf の値が高い場合は正しい単語に修正して採用する。低い場合はノイズとみなして排除する、という方針を採った。

表記の揺れ 今回は、重要語抽出の前処理として形態素解析しか行っていない。従って、表記の揺れについては、人手で統一をはかった。この作業は、同義

語辞書などを用意すれば、自動化することが可能になると思われる。

不要語の除去 品詞によるふり分けと tfidf によって重要でない高頻出の語（以下、不要語）はある程度自動的に除去される。それでもなお、抽出できない不要語も存在する。従って、人手による不要語の除去も行なった。

2.2 結果と考察

2.2.1 語の分類

以上のように抽出された語を、操作名と操作対象名に分類する³。形態素解析によって切り出された語には品詞のタグが付いているため、品詞による分類が可能である。ここでは、活動モデルの「操作名」に当たる語に「動詞」「名詞（サ変動詞語幹）」を、「操作対象名」に当たる語に「名詞（一般）」「名詞（形容動詞語幹）」⁴「未知語」を充てることにした。しかし、実際には、この振り分けには適合しないものが存在する。例えば、「カット」「メール」「マーク」の品詞はすべて、形態素解析結果はサ変動詞語幹となる。しかし、これらの語が「デザイン」に関わるワーカーの求職票では、下記のように「操作対象名」として使用されている。

Web デザイン・制作から、イラスト・雑誌の
カット・ポスター制作等、いたします。キャラクターイラスト、ロゴ マーク など柔軟に
対処します。

上記の場合には、「カット」「メール」「マーク」をそれぞれ、「操作対象名」に分類した。

また、同じ語でも文脈によって操作名である場合と操作対象名である場合がある。

ソフトは、アセンブラでのハードウェア割り込みから C++によるシステム 設計、DB 設計 までやれます。

機械 設計 製図全般と各種 CAD 図面作成及びその他データ入力業務などを業務請負する
アウトソーシング専門社です。

³操作名と操作対象名の詳しい説明は [mas00] を参照

⁴形容動詞語幹を「操作対象名」とすることに違和感を覚えるかもしれないが、「簡易 文書」「精密 画像」などの連語を考えると、形容動詞を操作対象の属性としてとらえることができる。従って、ここではこの品詞に属する語も、操作対象名としてとらえる。

¹<http://www.fmj.net/sohoo/main.html/>

²<http://www.vision.cup.com/>

はじめの例は、ソフトウェア開発に従事する SOHO ワーカの求職票であるが、「設計」というサ変動詞語幹が、操作名として使用されている。次の例は、CAD オペレーティングに従事するワーカの記述である。ここで現れる「設計」は操作ではなく、CAD オペレータが操作対象として扱う図面の種類を指す。このような語は操作対象名として分類した。

2.2.2 各業種ごとの語彙についての考察

以上のようにして抽出した語の集合を各業務における活動を表現する語彙とする。ここでは、「ワープロ入力」、「Web 制作」の2業種を表現する語彙についての考察を行なう。

ワープロ入力 表1は、ワープロ入力関連の業務を表現する語彙の一覧である。操作名には「経理」や「翻訳」など、この業務に直接関係の無さそうな語も見られる。これは、ワープロ入力という業務が他の業務に比べて、それほど専門的なスキルを必要としないためと思われる。他の業務に携わるワーカが併せて請け負うと記述しているケースが多く見られるからである。

この傾向は、「初心者」「未熟」「未経験」など、他の業務では決して見られないような、語が現れていることから推測できる。

Web 製作 表2は、Web 製作関連の業務を表現する語彙の一覧である。この業務に特徴的なのは「シンプル」「クール」「インタラクティブ」など、製作物の特徴を表現する形容動詞語幹が多く見られることである。これは、成果物である Web ページの要求特性として、分野や目的とは違う、いわばイメージに関するも

表 1: 「ワープロ入力」の語彙

操作名	計算 翻訳 経理 校正 入力
操作対象名	建築 データ 文書 表 事務 テープ データベース 責任 勉強 厳守 希望 大丈夫 迅速 確実 丁寧 正確 まじめ 自信 初心者 実務 未熟 未経験 資格 簡単 大好き 好き 必要 大量 十分 急 豊富 詳細 テクニカル いろいろ ワープロ メール Excel Word CAD Access EXCEL DTP WORD Windows PC

表 2: 「Web 製作」の語彙

操作名	デザイン 構築 運営 更新 プログラミング
操作対象名	広告 データ イラスト ページ データベース アプリケーション 画像 文書 ネットワーク コンテンツ 自信 正確 詳細 確実 勉強 大丈夫 丁寧 勤務 簡単 希望 必要 シンプル 完全 迅速 格安 ビジュアル クール 独自 安価 ショップ 様々 小規模 色々 インタラクティブ 効果 ブラウザ ワープロ メール 言語 技術 HP Web CGI HTML DTP CG Excel Access Windows VB Photoshop Flash Word WEB

のが存在するからであると考えられる。また、「安価」「格安」などの価格を表す語も、他の業務ではあまり見られない。

3 記述項目の抽出

3.1 手続き

前章では、ワープロ入力に従事する SOHO ワーカと Web 制作に従事する SOHO ワーカの求職票に現れる高頻度語についてそれぞれ考察を行なった。その結果、各業種に特徴的に出現する語彙の存在が確認された。この章ではすべての業種に共通する記述の傾向、すなわち SOHO ワーカの求職票の記述スキーマについて考察を行う。このため、まず抽出された語彙を分類し、それらの項目の性質を検討する。

図1は、前章の重要語にもとづき求職票の表現項目を分類したものである。下記に、各項目の概要を述べる。

【Operation】 この項目は操作に関する記述を示す。例えば「編集」「入力」「デザイン」などである。

【Object】 この項目は特に業務の成果物に関する記述を示す。例えば、「作品」「文書」「データ」「表」など成果物の形態や「イラスト」「ロゴ」「音」「コンテンツ」など、成果物の素材となるもの、あるいは成果物そのものなどが含まれる。

【Domain】 この項目は、主に Object の属する分野に関する記述を示す。例えば、「ビジネス」「テクニカル」などは Object 「文書」の分野を表現す。また、「音

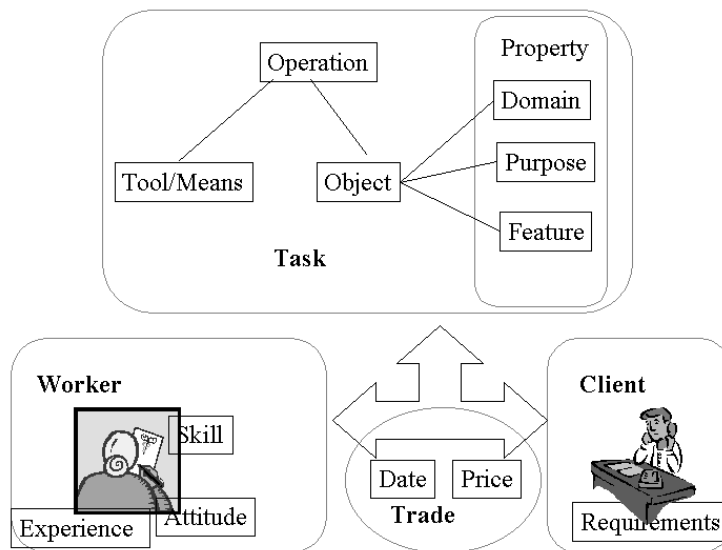


図 1: SOHO ワーカの求職票に見られる記述内容の分類

楽」「健康」などは、Object「記事」の属する分野を表現している。

【Purpose】この項目は、主に Object の目的や用途を表現する記述を示す。例えば、「配管」「建築」などは、Object「図面」の目的に関する語である。また、「CM」「イベント」「カラオケ」などは Object「音楽」の用途を表現する語である。

【Feature】この項目は、主に Object の特徴を表現する記述を示す。例えば、「cool」「可愛い」などは Object「イラスト」の特徴に関する語である。また、「詳細」「大量」などは Object「文書」の特徴を表現する語である。

【Attitude】この項目は、主に Worker の業務に対する態度や姿勢に関する表現を示す。例えば、「正確」「丁寧」「確実」などの語がこの項目に含まれる。

【Skill】この項目は主に Worker の能力や資格に関する表現を示す。例えば、「TOEIC」「会計士」「留学」などの語がこの項目に含まれる。

【Experience】この項目は主に worker の経験に関する表現を示す。「勤務」「実務」「従事」などの語がこの項目に含まれる。

【Requiemment】この項目は、発注者側の要求に関する記述を示す⁵。例えば、「希望」「要求」「気軽」などの語がこの項目に含まれる。

⁵実際の要求ではなく、あくまでワーカがクライアントの立場に立った上での想定である点に留意されたい

【Tool/Means】この項目は、主に業務を遂行するための、道具や手段に関する記述を示すものである。例えば、「Java 言語」「Access」「オブジェクト指向」などは「ソフトウェア開発・SE」の業務における手段を表現する語である。

4 数量化 II 類による業種間の差異の抽出

4.1 量的分析の試み

前々章では、SOHO ワーカの求職票を構成する語彙の各業務にみられる性質について考察を行った。さらに、前章ではすべての業務に共通する記述項目について、検討を行なった。この章では、それらの記述項目のうち、各業種においてどの項目が特徴的に現れるのかについての定量的な分析を試みる。

ここでは、各業務の求職票にどのような項目が記述されているかを、数量化 II 類⁶を適用した分析を行なう。これによって、各業務の求職表の記述にどの項目が主要になっているかが明らかになる。この結果から業種間にどのような差異があるかを考察する。

⁶数量化 II 類については [wat88] [wat92] 参照。

4.2 手続き

「翻訳」、「ソフトウェア開発・SE」、「経理・会計」の3業種を分析対象とした。各業種それぞれ30件の求職票を使用した。

求職票は統計分析にかけるため、数値データにコード化した。このデータの形式は、図3にある項目名⁷、すなわち *operation, object, *domain/purpose, feature, attitude, *skill/experience, requiement, price, date, *tool/means* を説明変数、業種 (*business-category*) を従属変数とする二元表である。

格安の料金 で 起票から決算 まで経理事務代行致します。一人で仕事をしているので 経理業務まで手が回らないとお悩み の事業主の方は是非御連絡下さい。

例えば、上記のような経理・会計に関する求職票を想定する。すべての変数の初期値は0とする。まず変数 *business-category* に業種を示すコードを設定する。次に、記述内容から、各項目に該当する箇所があれば、それに対応する変数が1となる。例えば、ここでは「起票」「決算」「代行」などの操作名が含まれているので、変数 *operation* が1に、さらに、「経理業務まで手が回らないとお悩み」はワーカがクライアントの要求に言及しているときのみ、変数 *requiement* が1になる。さらに、「格安の料金」という箇所は、取り引きの価格に関する記述なので、変数 *price* が1になる。

4.3 業種間の差異の検討

上記のような手続きで行なった数量化 II 類分析結果を元に、各業種における重要項目について考察を行なう。

「経理・事務」業務を説明するカテゴリースコアグラフである。この業種を説明するのに主要な項目は、「発注者の要求」(偏相関 0.38)、「操作名」(偏相関 0.34)、「価格」(偏相関 0.27) であることが分かる。一方、あまり重要でない項目(無、となっている項目)が「道具・手段」であることも分かる。

操作対象に関する項目ではなく操作名や要求が主要因として抽出されたのは、この業務の成果として期待

カテゴリースコア

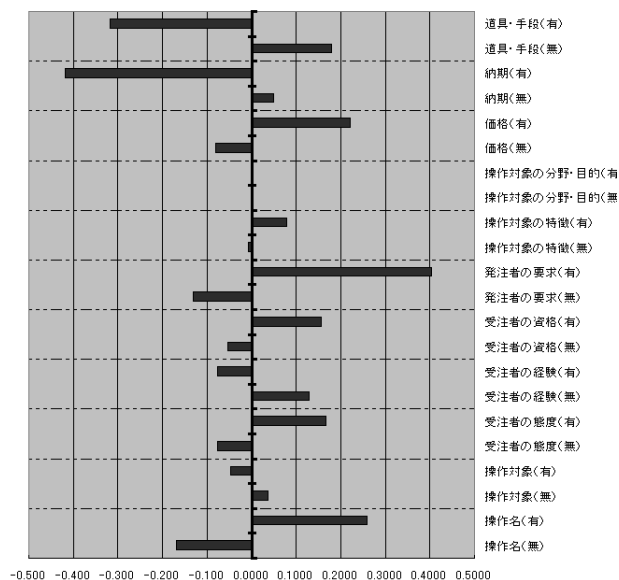


図 2: 「経理・事務」業務のカテゴリースコアグラフ

されるのが具象物ではなく、むしろ無形のサービスが主であるためだと思われる。

図3は「翻訳」業務を説明するカテゴリースコアグラフである。この業種を説明するのに主要な項目は、「期日」(偏相関 0.35)、「操作対象の分野・目的」(偏相関 0.14) であることが分かる。一方、あまり重要でない項目が「道具・手段」「要求」(偏相関)「操作名」「ワーカの態度」であることも分かる。

期日と操作対象の分野・目的が主要因として抽出されたのは、翻訳業務において最も重要なのはその翻訳対象の領域であるためであろう。逆に、それ以外の要素、例えば、道具や手段、相手の要求などはそれほど考慮に入れなくてもいい事柄であるからだと思われる。また、操作名に関しても、翻訳業務の場合、「翻訳」以外の操作が特にないため、あまり重要でない項目となっている。

図4は「ソフト開発・SE」業務を説明するカテゴリースコアグラフである。この業種を説明するのに主要な項目は、突出して「道具・手段」(偏相関 0.76)、次に「納期」(偏相関 0.17)、であることが分かる。

この業務では、プログラミング言語の種類とソフトウェアの動作するプラットフォームの種類が最も重要である。逆にそれ以外の項目、例えばクライアントの要求やソフトの適用領域などについては、それほど記

⁷*印は複数の項目を統合したもの

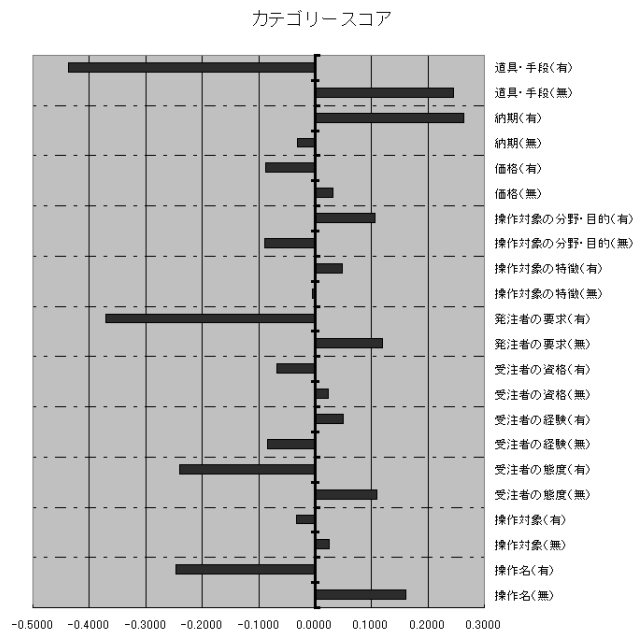


図 3: 「翻訳」業務のカテゴリースコアグラフ

述する必要を感じていないと思われる。

全体的に、偏相関係数がそれほど高くないが、これには二つの要因が考えられる。まず、ここで比較したのは、3業種であったが、これらの業種自体の差がそれほど顕著でなかった可能性である。次に、説明変数が適切でなかった可能性である。まず、前者については対象とする業種を増やすことで、検証できる。また、後者については、より適切な説明変数の設定を検討することが課題となる。

5 今後の課題

今回の調査は、なるべく調査者の主観を排除するため、分析のための一連の手続きを自動的もしくは形式的なものにするよう努めた。今後、この一連の手続きをどこまで自動化できるかが今後の大きな課題の一つである。

さらに、今回の分析は、SOHO ワーカーが受発注関係の成立にどのような項目を重要ととらえているかを探ることが目的であった。もっとも、今回の手法は、自然言語で記述された他の対象への適用も考えられる。例えば、自由記述形式のアンケートなどの分析に、今回の手続きを用いることも考えられる。

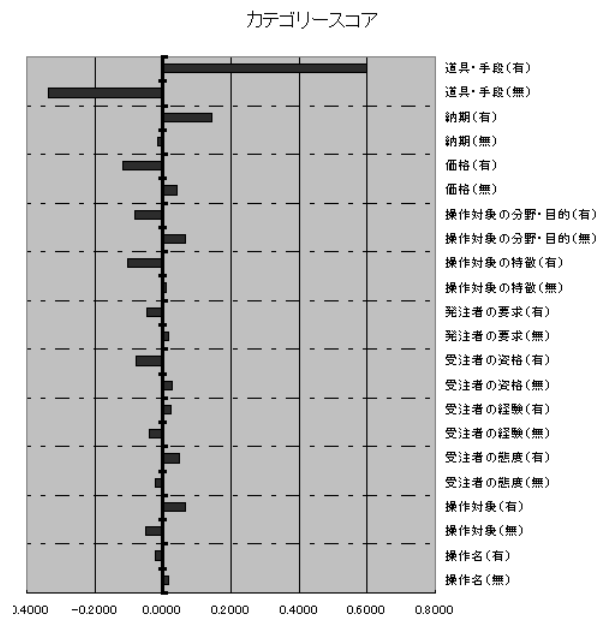


図 4: 「ソフトウェア開発・SE」のカテゴリースコアグラフ

謝辞

本研究に関して調査にご協力いただいた、SohoolJapan 様、クリエイターズゲート様、有益なコメントをいただいた富士ゼロックスキャリアネット (株) の皆様、日頃から議論している知識研究室 BKR グループの各メンバーに感謝の意を表します。

参考文献

- [Mar97] T.W.Malone 他. Tools for inventing oragnizations: Towared a handbook of organizational processes. <http://ccs.mit.edu.CCSWP198/>, 1997.
- [mas00] Y.Masuda 他. Towards the cross-organizational work process coordination and enactment. *Proceedings of AAAI Spring Symposium Series 2000 Bringing Knowledge to Business Process*, 2000.
- [mat99] 松本裕治他. 日本語形態素解析システム「茶筌」 version 2.0 使用説明書第 2 版. 1999.
- [wat88] 渡部洋. 心理・教育のための多変量解析入門——基礎編. 福村書店, 1988.
- [wat92] 渡部洋. 心理・教育のための多変量解析入門——事例編. 福村書店, 1992.