

クラウドソーシング発注文書におけるレジビリティの量的分析

著者	岩崎 拓也, 井上 雄太
雑誌名	言語資源活用ワークショップ発表論文集
巻	3
ページ	399-414
発行年	2018
URL	http://doi.org/10.15084/00001674

クラウドソーシング発注文書におけるレジビリティの量的分析

岩崎 拓也 (国立国語研究所理論・対照研究領域／一橋大学大学院言語社会研究科) †

井上 雄太 (一橋大学大学院言語社会研究科) ††

Quantitative Analysis of Legibility in Crowdsourcing Purchase Orders

Takuya Iwasaki (NINJAL / Hitotsubashi University Graduate School of Language and Society)

Yuta Inoue (Hitotsubashi University Graduate School of Language and Society)

要旨

インターネットを介して業務を遂行するクラウドソーシングにおいて、まず重要になるのが可読性である。可読性には、文法の難易度にかかわるリーダビリティと文章の見やすさにかかわるレジビリティがある。可読性を上げるためには、わかりやすい文法や表現を使用するだけでなく、見やすい文書にしなければならない。レジビリティを上げる要素として考えられるのは、句読点やカッコといった記号、文と文の間の空白行の挿入などであるが、これらについて計量的な観点からの分析は多くない。本研究では、クラウドソーシングで実際に使われた発注文書のデータベースを元に、記号や空白行に焦点を当て、その多寡と応募数との相関の分析を行った。その結果、記号は hourly 型の支払い形式、改行は hourly 型と task 型の支払い形式、空白行は task 型の支払い形式の間において弱い相関が認められた。

1. はじめに

クラウドソーシングは、「高速」「低コスト」が利点ではあるものの、ワーカー(受注者)への対応の難しさがある。受注者を段階的に育成する方法も提案されているが(芦川・川村・大須賀 2017)、単発的な発注である場合は育成ができず、効果が得られにくいことも考えられる。そこで、重要になるのは発注文書である。内容が明瞭でわかりやすい発注文書であれば、より多くのワーカーが応募してくる可能性が高くなると考えられる。実際に、清水・中川(2015)では、発注文書内の質問が曖昧であるという問題を挙げている。このことは、発注者に対してどうすれば文書がわかりやすくなるかという説明が求められていることを示している。本研究では、クラウドソーシングの発注文書の本文(以下、発注文書)を対象として分析を行う。クラウドソーシングにおいて、ワーカーは Web サイトにアクセスした後、タイトル一覧を読み、自らが得意な(または作業したい)案件をクリックし、発注文書を読む。これを繰り返し行い、最も自分に合う案件に応募する。この際、本文の内容が見にくく、内容がわかりにくい場合は応募する可能性が低くなると考えられる。通常、発注文書を読みやすくする要素としては、簡潔な表現を用いることや内容を箇条書きにすることなどが挙げられる。また、箇条書きにする際や言及している内容を階層化する際に、記号の使用や空白行の挿入も行われる。このような要素は、いずれも書き手(クラウドソーシングにおいては発注者)が読み手(受注者)に対して、より目に止まるように見せたり、何らかの注釈を示したりするために使用されているものだと考えられる。実際に、有光ほか

† i.was.aki[at]hotmail.co.jp

†† inoue.yta[at]gmail.com

(2014) では、かぎカッコによる効果の検証を調査しており、調査の結果、読み手はほとんどの文においてかぎカッコがある文の方がわかりやすい文章であると感じていることを明らかにしている。しかしながら、記号による文字装飾の有無と応募数の相関、発注文書で使われる記号の種類とその使用傾向については計量的な立場からの分析は行われておらず、その実態は明らかにされていない。

2. 先行研究

クラウドソーシング発注文書データベースの構築については、井上 (2018) が初期段階での報告を行っている。井上 (2018) では、形態素解析を行ったデータベースにおける基礎統計量 (タイトルの形態素数と品詞別頻度、本文と件名の形態素数と品詞別頻度) について報告している。また、クラウドソーシングにおける文書の見やすさについての先行研究として、岩崎 (2018)、佐野 (2018) がある。岩崎 (2018) では、構築中のクラウドソーシング発注文書データベースにおける発注文書の依頼タイトルに焦点をあて、その中で使用されている記号の分析を行っている。その結果、タイトルの 94.6%において、何らかの記号が使用されていることが報告されている。また、記号の使用には、年代差、性差などの偏りも存在している可能性が示唆された。佐野 (2018) では、1000 件の発注文書内にある見出しを、使用されている記号を手がかりに分析している。その結果、見出しにおいては、記号の使用頻度が高く、その中でも墨付カッコ (【 】) の使用が多く、見た目にインパクトを与えていることを指摘している。

3. 本研究の目的

本研究では、上述した先行研究を踏まえ、「クラウドソーシング発注文書データベース (仮称)」(以下、CSDB) の構築方法の詳細を説明し、データの特徴に沿った傾向を挙げる。その上で、次の 3 点を明らかにすることを試みる。

発注文書において、

- (1) 使用されている記号の種類と多寡。
- (2) 記号の多寡と応募数の相関。
- (3) 空白行の挿入と応募数の相関。

これらが明らかになれば、発注文書の読みやすい (または見やすい) 書き方の傾向がわかり、発注文書と応募数の相関が明らかになれば、ワーカーが興味を持つ発注文書の一端が集合的に明らかになると考えられる。

4. クラウドソーシング発注文書データベースの構築 (井上)

以下では、CSDB の元となったデータの概要を取り上げた後、DB の構築方法の詳細と仕様・本文の形態素解析結果の基本統計量を用い DB の全体像について解説する。

4.1 元となるデータについて

以下では、本研究で作成したデータベースの元となったデータの概要について解説する。データベースの元となったのは、日本最大級の総合型クラウドソーシングサイトを運営するクラウドワークス社の発注データ 1 ヶ月分 (2017 年 8 月) 28,896 件である。クラウドワークス社より提供された発注データは、案件ごとに id が割り振られ、基礎的な情報として件名・本文、付加的な情報として支払い形式・案件カテゴリ・期間・予算・応募数・ユーザ

一による評価等が含まれている。

本研究ではクラウドソーシングにおける業務の流れを大きく分ける仕事形式・支払い形式に着目する。支払い形式はテーブル上にて以下の4カテゴリに分けられる。

(4) hourly 型：時間単価制。EC サイトの運営など長期間のものに多い。

(5) fixed_price 型：固定報酬制。記事制作など作業量の決まったものに多い。

(6) competition 型：応募採用者のみに支払われる形式。ロゴマークの募集など。

(7) task 型：固定報酬制。アンケート・レビューなど短時間で終わるものに多い。

また、これらの4カテゴリは、サイト上では(4)(5)がプロジェクト形式、(6)がコンペ形式、(7)がタスク形式という3つの仕事形式でまとめられている。それぞれの形式の発注件数は以下の表1のとおりであり、fixed_price 型と task 型が全体の96.2%を占めている。

表1 CSDBにおける仕事形式（支払い形式）別の発注件数とその割合

仕事形式	支払い形式	発注件数	割合
プロジェクト形式	hourly 型（時間単価制）	787	2.72%
	fixed_price 型（固定報酬制）	13897	48.09%
コンペ形式	competition 型	322	1.11%
タスク形式	task 型	13890	48.07%
合計	-	28896	-

4.2 データベースの構築方法

CSDB では提供された発注データと紐付ける形で、発注文の件名・本文に対し形態素解析¹を行った。形態素解析エンジンには MeCab (ver 0.996) を利用し、解析用辞書には unidic-cwj-2.2.0 を用いた。解析結果の整形には Python 3.6.4 を利用し、csv 形式にて出力した。

形態素解析の結果から、書字形・語彙素・語彙素読み・品詞（細分類まで）・文字種を抽出した後、書字形の文字数・一致用文字列を追加した。一致用文字列としては、語彙素・語彙素読み・品詞（細分類まで）を結合したものを採用した。これは既存の形態素解析による形態素 id のみでは、辛い（からい／つらい）最中（さいちゅう／もなか）といった同一の語彙素が語彙素読みによって複数の意味を持つ場合に対応できないことによる。

解析結果に対し、件名内の各形態素には形態素 id・案件 id を、本文内の各形態素には行 id²・形態素 id・案件 id を付与することで、CSDB 内にて各形態素を一意に呼び出すことが可能となっている。また、共通の案件 id を持つことにより解析結果を閲覧数などの提供データ内の付加情報と紐付けた分析も可能である。本文内の改行情報は、本文全体の末尾にあるもののみを削除し、空白行の情報を取り出しやすいよう配慮した。

なお、クラウドワークスの発注文作成画面では、本文内の文字列に対し文字サイズ・文字色・ボールドによる装飾を指定でき、これらは html タグによって記載される。提供された

¹ 正確には形態素ではなく短単位による解析であるが、通例に従い以降の文章では短単位での解析を形態素解析として扱う。

² 文単位ではなく行単位に id を付けたことは、対象となる本文に箇条書きなど末尾がないものが多数出現することによる。当 DB における行単位は、改行コードによって区切られたものとし、表示環境に大きく依存する自動的な文字送りに関しては対象としない。

データ内にて本文に html タグを含む発注文は 18,433 件, 含まない発注文は 10,463 件であった。装飾用の html タグは形態素解析の際に誤解析の元となるため, 前処理の段階で削除した。html による文字装飾を CSDB に組み込む点に関しては, 今後の課題とした。また, html タグ付きの本文に関しては, ブラウザ表示時には適用されない改行タグ以外の改行情報を削除し, 改行タグのみを改行情報として認定した。前処理段階にて改行は LF (¥n) に, 文字コードは UTF-8 に統一した。

4.3 データベース上における基本統計量

以下では, CSDB 内での発注文本文の支払い形式別の形態素数について概観する。まず, 発注文書における形態素解析による品詞大分類ごとの総数を形式別に示したものを表 2 に示す。また, 表 3 は発注文書 1 件ごとの形態素数の平均である。

表 2 CSDB における支払い形式別の形態素数 (品詞大分類)

品詞大分類	hourly 型	fixed_price 型	competition 型	task 型	全体
改行	19311	815439	20472	478717	1333939
感動詞	237	8929	145	3327	12638
記号	4039	110774	11661	83346	209820
空白	1705	88203	3200	51644	144752
形状詞	1962	87455	2460	50304	142181
形容詞	995	43315	1497	30289	76096
助詞	31374	1459333	39529	1002781	2533017
助動詞	10005	533388	13054	370627	927074
接続詞	212	12257	395	9570	22434
接頭辞	3142	216411	3845	110514	333912
接尾辞	4168	154069	5171	96417	259825
代名詞	692	28970	648	16731	47041
動詞	17612	830531	21472	578518	1448133
副詞	820	49007	1513	33564	84904
補助記号	29636	1160583	37991	868297	2096507
名詞	67027	2570670	79549	1841809	4559055
連体詞	609	25805	670	19638	46722
合計	193546	8195139	243272	5646093	14278050

表3 発注文書1件あたりの平均形態素数（支払い形式別）

品詞大分類	hourly 型	fixed_price 型	competition 型	task 型	全体
改行	24.54	58.68	63.58	34.46	46.16
感動詞	0.3	0.64	0.45	0.24	0.44
記号	5.13	7.97	36.21	6	7.26
空白	2.17	6.35	9.94	3.72	5.01
形状詞	2.49	6.29	7.64	3.62	4.92
形容詞	1.26	3.12	4.65	2.18	2.63
助詞	39.87	105.01	122.76	72.19	87.66
助動詞	12.71	38.38	40.54	26.68	32.08
接続詞	0.27	0.88	1.23	0.69	0.78
接頭辞	3.99	15.57	11.94	7.96	11.56
接尾辞	5.3	11.09	16.06	6.94	8.99
代名詞	0.88	2.08	2.01	1.2	1.63
動詞	22.38	59.76	66.68	41.65	50.12
副詞	1.04	3.53	4.7	2.42	2.94
補助記号	37.66	83.51	117.98	62.51	72.55
名詞	85.17	184.98	247.05	132.6	157.77
連体詞	0.77	1.86	2.08	1.41	1.62
合計	245.93	589.71	755.5	406.49	494.12

次に、発注文書1件あたりの改行を含む形態素数の基本統計量を表4に示す。4.1で確認したように、作業内容に対する指示を比較的多く必要としない task 型・competition 型は、形態素数が少ない値で分布している。また、発注後のやりとりをすることが可能なため、業務内容を発注文の中でどこまで説明するかが発注者次第となる hourly 型・fixed_price 型では、形態素数のばらつきが大きいことが見て取れる。

表4 発注文書1件あたりの形態素数（支払い形式別）

	hourly 型	fixed_price 型	competition 型	task 型	全体
Max.	1589	2297	1758	2220	2297
3rd Qu.	913	807	374	571	724
Mean	601.1	589.7	309.1	406.5	494.1
Median	501	562	276	353	438
1st Qu.	296.5	319	144	191	248
Min.	4	1	2	3	1
SD	381.1	333.1	242.3	280.4	321.7
Var	145233.9	110980.6	58721.1	78598.4	103508.5

5. クラウドソーシング発注文書データベースにおける記号と空白行（岩崎）

以下では、CSDBにおいて使用されている記号の種類と多寡をそれぞれ取り上げ、その傾向を明らかにする。また、空白行の挿入が1文書あたりどのくらいの割合で挿入されているかといった傾向についても明らかにする。

5.1 分析対象

クラウドワークスにおいて、仕事の依頼形式は複数存在している。今回は、「プロジェクト形式 (hourly 型), (fixed_price 型)」と「コンペ形式 (competition 型)」、「タスク形式 (task 型) 全ての仕事形式を対象とし、支払い形式ごとにどのように記号と改行、空白行が挿入されているかを分析していく。

5.2 発注文書における記号の種類と多寡

以下の表5に各支払い形式において使用されていた記号とその使用頻度（上位10語）を挙げる。それぞれの型に出現した全記号とその頻度については、本論文の最後に示す。まず、使用されている記号の種類を見てみると、hourly型においては70種類、fixed_price型においては155種類、competition型においては101種類、task型においては134種類の記号の使用が確認された。同データベースを使用した発注文書のタイトルにおいては、121種類の記号が使用されており（岩崎 2018）、発注文書においてはタイトルよりも記号の種類が多いことがわかった。なお、それぞれの形式に使用されていた記号の総頻度は、hourly型が29,636、fixed_price型が1,160,554、competition型が37,991、task型は868,297であった。

表5 発注文書で使用されていた記号の種類と頻度（上位10語）

	hourly 型			fixed_price 型			competition 型			task 型		
	語彙素	頻度	割合	語彙素	頻度	割合	語彙素	頻度	割合	語彙素	頻度	割合
1	、	5178	17.47%	。	209953	18.09%	。	5808	15.29%	。	164034	18.89%
2	・	4825	16.28%	、	194088	16.72%	、	5774	15.20%	、	128230	14.77%
3	。	4101	13.84%	・	148402	12.79%	・	4419	11.63%	・	103096	11.87%
4)	1838	6.20%)	52860	4.55%	/	3267	8.60%	※	38618	4.45%
5	(1823	6.15%	(51431	4.43%	.	2298	6.05%	=	36925	4.25%
6	/	1354	4.57%	—	42895	3.70%	:	1959	5.16%	【	34057	3.92%
7	:	1266	4.27%	!	37845	3.26%)	1889	4.97%	】	34042	3.92%
8	【	1082	3.65%	:	36201	3.12%	(1887	4.97%)	32696	3.77%
9	】	1082	3.65%	■	32294	2.78%	▽	1188	3.13%	(31561	3.63%
10	～	946	3.19%	/	31673	2.73%	※	1181	3.11%	/	27378	3.15%

#網掛け箇所は、全ての形式で10位以内だった記号である。

(8) ☆☆☆こんな方は必見☆☆☆

- ・面白い記事を読むのが好きな人
- ・1日2～3時間自由な時間がある人
- ・ネットサーフィンが好きな人

(job_offer_id: 1547354)

(9) [形式]

以下の記事のように小見出しと段落を付けてください

<http://tsuufuu.com/gout-woman>

[掲載サイトと参考サイト]

<http://tsuufuu.com/gout-obesity-urolith>

<http://tsuufuu.com/gout-measures>

(job_offer_id: 1725117)

次に、使用されていた記号について見ていく。全ての形式において使用されていた記号は、句読点と中点（・）、カッコ、スラッシュ（/）であった。中でも、句読点と中点（・）が多く使用されていることがわかる。これらの記号のうち、句読点は文を書く際に必須の記号である。また、中点は（8）のように箇条書きをする際の頭に打たれることが多く、発注文書を書く際に見やすさとわかりやすさを工夫する上で用いられる記号であると言える。スラッシュは（9）のように成果物のイメージや参考となる記事やサイトの URL の一部であると考えられる。文書による伝達を行う必要があるクラウドソーシングにおいて、発注者は受注者に発注者が求める成果物を URL のリンク先を介してあらかじめ提示している。これらの記号は、いずれもクラウドソーシングの発注文書において大きな役割を果たしていると考えられる。

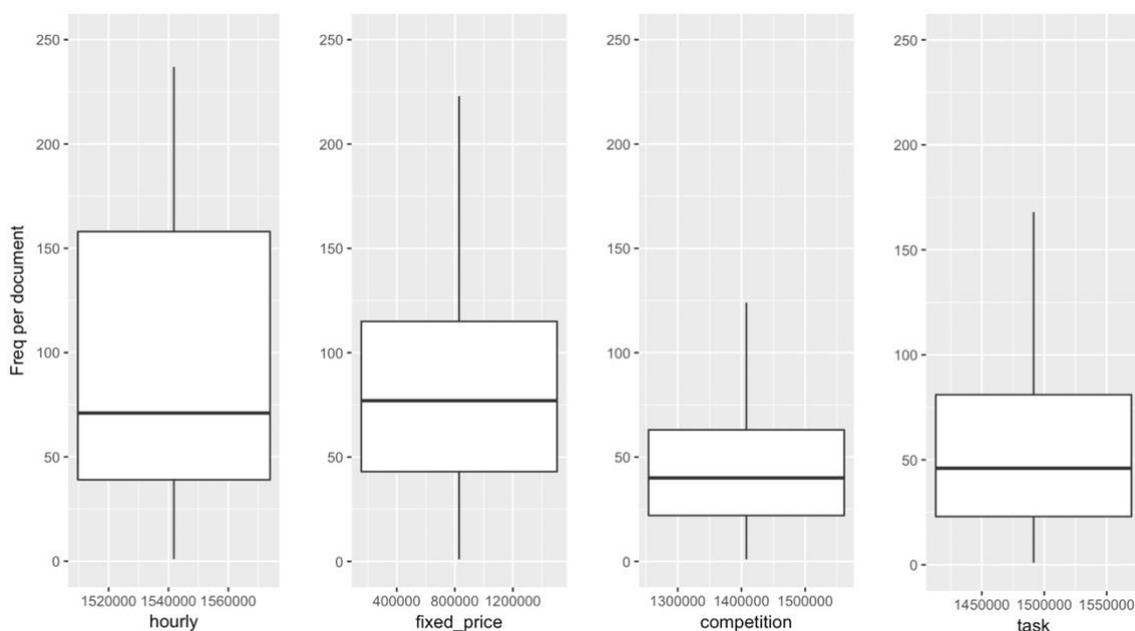


図 1 発注文書 1 件あたりの記号の数

表 6 発注文書 1 件あたりの記号の数

	hourly 型	fixed_price 型	competition 型	task 型
Max.	237	604	386	755
3rd Qu.	158	115	63	82
Mean	92.32	83.65	48.33	62.57
Median	71	77	40.5	47
1st Qu.	39	44	22	23
Min.	1	1	1	1
SD	64.7	52.52	38.74	56.33
Var	4186.19	2758.66	1500.42	3172.92

図 1 と表 6 は、各仕事形式の発注文書 1 件あたりの記号の使用頻度をまとめた boxplot と基本統計量である³。図 1 を見ると、hourly 型の分散が非常に大きく、同じ「プロジェクト形式」であっても、支払い方法の違いにより、記号の使用量に差があることがわかる。また、competition 型は分散が非常に小さく、平均も他の形式に比べて小さいことがわかる。「コンペ形式」は、デザイン系の発注が多く、発注文書 1 件あたりの平均形態素数も competition 型が最も少なく（competition 型:309.1, hourly 型:601.1, fixed_price 型:589.7, task 型:406.5）、competition 型の発注文書は最低限の指定しか書いていないことが多いと考えられる。そのため、記号の使用も少なくなっていると考えられる。

各形式における記号の使用と応募数（クラウドワークスでは「提案人数」と表記）に相関があるかを分析した。その結果、hourly 型にのみ弱い相関が認められた ($r=0.2550893$)。

5.3 発注文書の長さ（改行数）

次に、発注文書の長さについて見ていく。ここでの長さは文書において改行した回数を指す。通常、クラウドソーシングにおける発注文書は、Web ページ上で閲覧するものである。Web ページ上では一度に表示される文章量が限られているため、文章が多く改行されている場合はそれだけ多くページをスクロールしなければならない。しかし、文章が短すぎると、十分な説明を行うことが難しくなり、受注者（読み手）にとっては、わかりにくい（応募しにくい）発注文書となってしまう。

³ boxplot の外れ値は除去して表示している（以下、全ての boxplot も同様である）。

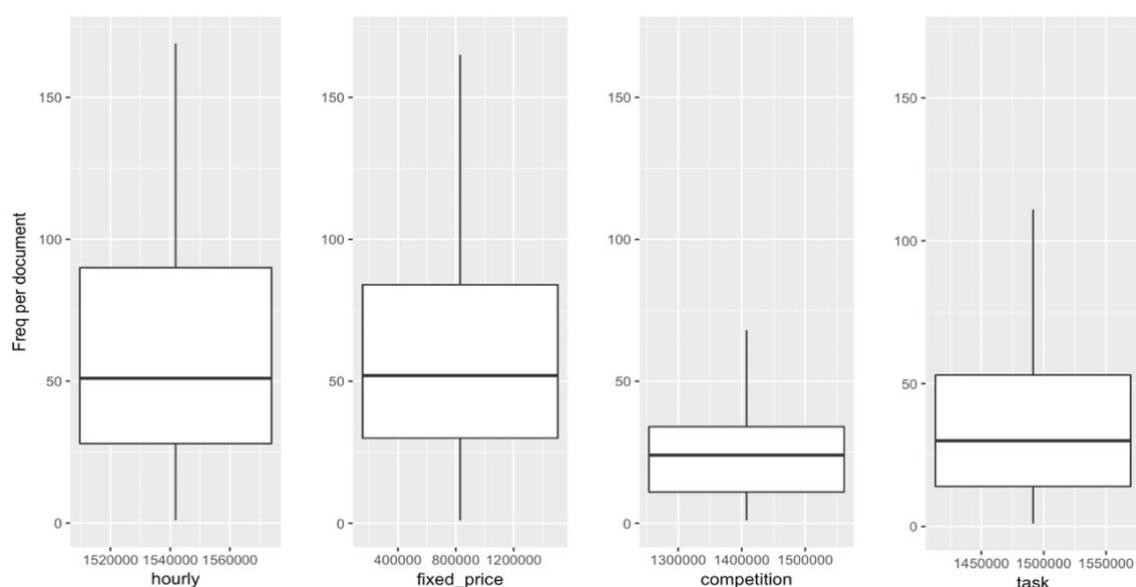


図2 発注文書1件あたりの長さ(改行数)

表7 発注文書1件あたりの長さ(改行数)

	hourly 型	fixed_price 型	competition 型	task 型
Max.	169	324	142	187
3rd Qu.	90	86	34	53
Mean	60.97	59.68	27	35
Median	51	53	24	30
1st Qu.	28	30	22	14
Min.	1	1	1	1
SD	41.36	37.28	20.74	26.04
Var	1710.88	1390.17	429.97	677.94

図2と表7は、各形式の発注文書1件ごとの長さ(改行数)をまとめたboxplotと基本統計量である。前節で見た記号とは異なり、「プロジェクト形式」であるhourly型とfixed_price型は類似した分布であることがわかる。competition型とtask型については、改行が少なく短い発注文書がほとんどであることがわかる。文章量という視点から分析を行う場合は、支払い形式ではなく、仕事形式の単位で分析を行うことが望ましいと考えられる。

(10) 先月の実績をご報告下さい。

(job_offer_id: 1507893)

(11) エクセルファイルの件です

(job_offer_id: 1558011)

また、表7を見ると、最小値が1となっているが、これは(10)や(11)のように本文が1行のみの発注文書である。ただし、このような例は、前回の発注の続きであったり、特定の受注者に対する発注であったりするなど、通常の発注とは異なる発注であると考えら

れる。各形式における改行の使用と応募数に相関があるかを分析した。その結果、hourly 型 ($r=0.2227566$) と task 型 ($r=0.2316384$) に弱い相関が認められた。

5.4 発注文書における空白行の数

最後に、発注文書内で挿入された空白行について分析を行う。空白行が挿入されている発注文書を計算したところ、hourly 型 322 件中 311 件 (93.48%)、fixed_price 型 13897 件中 13608 件 (97.92%)、competition 型 787 件中 734 件 (93.48%)、task 型 13890 件中 12984 件 (93.48%) と全ての形式で 90%以上の割合で空白行が挿入されていた。次の図 3 は、発注文書 1 件あたりの空白行の数をまとめた boxplot である。記号や改行の数と比較すると、hourly 型において今までの傾向とは異なり、分散が小さいことがわかる。一方で、同じ「プロジェクト形式」の fixed_price 型は hourly 型より分散が大きい。つまり、hourly 型では空白行の挿入に一定の傾向が存在する可能性がある。各形式における空白行の使用と応募数に相関があるかを分析した。その結果、task 型にのみ弱い相関が認められた ($r=0.211738$)。

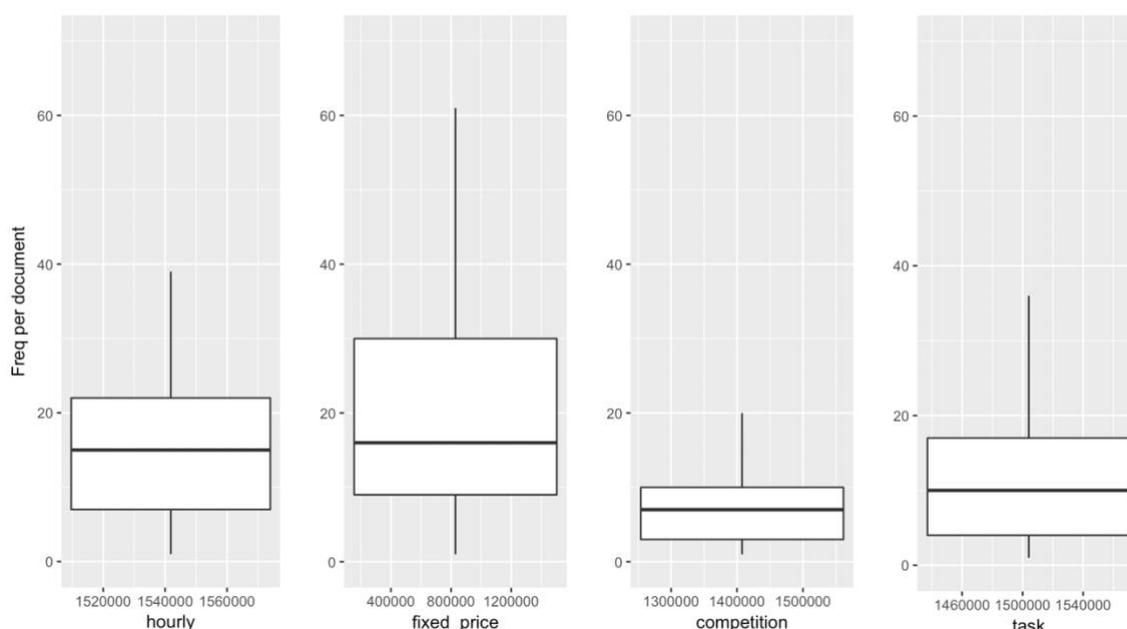


図 3 発注文書 1 件あたりの空白行の数

表 8 発注文書 1 件あたりの空白行の数

	hourly 型	fixed_price 型	competition 型	task 型
Max.	75	225	71	99
3rd Qu.	22	32.25	10	17
Mean	18.91	22.82	8.101	12.23
Median	15	16	7	10
1st Qu.	7	9	3	4
Min.	1	1	1	1
SD	16.32	20.4	6.97	10.46
Var	16.32	416.03	48.55	109.38

6. まとめと今後の課題（井上・岩崎）

以上、本研究では、インターネット上において「クラウドソーシング発注文書データベース（仮称）」の構築方法の詳細を説明し、データの特徴に沿った傾向を挙げた。その上で、発注文書において、使用されている記号の種類と多寡、発注文書の長さや空白行の挿入数を調査し、各変数と応募数の相関を見た。その結果、記号は hourly 型、改行は hourly 型と task 型、空白行は task 型との間において、それぞれ弱い相関が認められた。

今後の課題を以下に挙げる。現時点での CSDB はクラウドワークス社の 1 ヶ月分の発注データを用いて構築されているが、将来的に大規模化が予定されている。これに伴うデータの肥大化と処理時の付加の増大に対応するため、本文の現在形態素解析結果のテーブルを分割し、他のデータと連携しやすい形の間テーブルを設計する必要がある。また、発注文や件名と応募数・評価などとの相関を分析する際の精度を高めるため、単価・期間・サイト内での案件を目立たせるために使用されるオプション機能の有無などを加味して標準化したコアデータの作成についても検討する必要がある。

今回は発注文書における記号の種類と多寡のみを明らかにしたが、これらの記号が発注文書内でどのような機能を持っているかを分析する必要がある。また、空白行については、発注文書 1 件あたりの数のみを明らかにしたが、実際に発注文書を読んでもみると、連続して空白行が挿入される場合もある。そのため、どのぐらいの空白行が連続して挿入されるのかといった点も考慮する必要がある。また、空白行の挿入は、書き手の段落意識と関係があると思われる。この点についても分析して行く必要がある。

謝 辞

本研究は、国立国語研究所機関拠点型基幹研究プロジェクト「日本語学習者のコミュニケーションの多角的解明」（プロジェクトリーダー：石黒圭）の研究成果の一部であり、JSPS 科研費 17K18504 挑戦的研究（萌芽）助成を受けたものである。

文 献

- 有光隆・八木秀次・呉志強・李在勲（2014）「読みやすいテキストと日本語の文章構造に関する一考察」、『工学教育』, 62:2, pp. 2_51-2_56.
- 芦川将之・川村隆浩・大須賀昭彦（2017）「クラウドソーシングワーカーの段階的育成方法の提案」、『人工知能学会論文誌』, 32:3, pp. B-G81_1-13.
- 井上雄太（2018）「発注文書の分析に有効なデータベースの構築」『第 20 回日本テレワーク学会研究発表大会』発表資料, 千葉商科大学, 2018 年 7 月 8 日.
- 岩崎拓也（2018）「発注文書で目を惹く記号・顔文字の使い方」『第 20 回日本テレワーク学会研究発表大会予稿集』, pp.169-172.
- 佐野彩子（2018）「内容が一目でわかる発注文書の見出し」『第 20 回日本テレワーク学会研究発表大会』発表資料, 千葉商科大学, 2018 年 7 月 8 日.
- 清水伸幸・中川雅史（2015）「クラウドソーシングの現状と可能性：2. マイクロタスク型クラウドソーシングの現状と課題 -実際の運用の知見から-」, 『情報処理』, 56:9 pp.886-890.

資料1 hourly型の発注文書で使用されていた記号の種類と頻度

	語彙素	頻度	割合 (%)
1	、	5178	17.47%
2	・	4825	16.28%
3	。	4101	13.84%
4)	1838	6.20%
5	(1823	6.15%
6	/	1354	4.57%
7	:	1266	4.27%
8	【	1082	3.65%
9	】	1082	3.65%
10	～	946	3.19%
11	●	882	2.98%
12	.	617	2.08%
13	※	544	1.84%
14	!	477	1.61%
15	■	375	1.27%
16	▽	330	1.11%
17	「	321	1.08%
18	」	321	1.08%
19	=	267	0.90%
20	▼	264	0.89%
21	★	233	0.79%
22	—	143	0.48%
23	『	127	0.43%
24	』	127	0.43%
25	α	114	0.38%
26	—	90	0.30%
27	?	84	0.28%
28	◆	76	0.26%
29	&	73	0.25%
30	>	73	0.25%
31	<	72	0.24%
32	×	71	0.24%
33	~	45	0.15%
34	○	41	0.14%
35	♪	32	0.11%
36	[31	0.10%
37]	31	0.10%
38	...	29	0.10%
39	っ	26	0.09%
40	↓	25	0.08%
41	+	23	0.08%
42	☆	19	0.06%
43	→	15	0.05%
44	◇	15	0.05%
45	,	14	0.05%
46	*	13	0.04%
47	^	12	0.04%
48	/	11	0.04%
49	#	11	0.04%
50	%	8	0.03%
51	《	6	0.02%
52	》	6	0.02%
53	◎	6	0.02%
54	—	5	0.02%
55	@	5	0.02%
56	-	4	0.01%
57	⇒	4	0.01%
58	÷	4	0.01%
59	unk	4	0.01%
60	^ ^	2	0.01%
61	□	2	0.01%
62	⑥	2	0.01%
63	⑦	2	0.01%
64	∴	1	0.003%
65	°	1	0.003%
66	<U+2661>	1	0.003%
67	<U+2666>	1	0.003%
68	○	1	0.003%
69	○	1	0.003%
70	↗	1	0.003%

資料2 fixed_price型の発注文書で使用されていた記号の種類と頻度

	語彙素	頻度	割合 (%)
1	。	209953	18.09%
2	、	194088	16.72%
3	・	148402	12.79%
4)	52860	4.55%
5	(51431	4.43%
6	—	42895	3.70%
7	!	37845	3.26%
8	:	36201	3.12%
9	■	32294	2.78%
10	/	31673	2.73%
11	※	30002	2.59%
12	.	25207	2.17%
13	【	24474	2.11%
14	】	24469	2.11%
15	=	22090	1.90%
16	「	20633	1.78%
17	」	20587	1.77%
18	～	15631	1.35%
19	♪	11141	0.96%
20	☆	8780	0.76%
21	—	7082	0.61%
22	▽	7010	0.60%
23	*	6984	0.60%
24	★	6561	0.57%
25	●	6544	0.56%
26	?	6415	0.55%
27	○	6155	0.53%
28	▼	6148	0.53%
29	↓	4545	0.39%
30	◆	3943	0.34%
31	—	3019	0.26%
32	^	2978	0.26%

33	%	2942	0.25%
34	→	2925	0.25%
35	×	2638	0.23%
36	『	2481	0.21%
37	』	2474	0.21%
38	>	2254	0.19%
39	~	2228	0.19%
40	<	2101	0.18%
41]	2082	0.18%
42	[2081	0.18%
43	+	2058	0.18%
44	&	1987	0.17%
45	...	1846	0.16%
46	/	1844	0.16%
47	ˆ	1525	0.13%
48	—	1279	0.11%
49	⇒	1266	0.11%
50	○	1263	0.11%
51	@	1137	0.10%
52	◎	1037	0.09%
53	◯	848	0.07%
54	<U+2661>	781	0.07%
55	⑦	780	0.07%
56	□	731	0.06%
57	\	695	0.06%
58	②	692	0.06%
59	③	688	0.06%
60	,	534	0.05%
61	unk	527	0.05%
62	¥	488	0.04%
63	《	480	0.04%
64	》	478	0.04%
65	..	460	0.04%
66	—	407	0.04%
67	④	328	0.03%
68	”	325	0.03%
69	<U+2013>	294	0.03%
70	°	264	0.02%
71	↑	253	0.02%
72	;	251	0.02%
73	⑥	239	0.02%
74	-	236	0.02%
75	—	232	0.02%
76	#	225	0.02%
77	◇	200	0.02%
78	<U+266B>	187	0.02%
79		139	0.01%
80	△	107	0.01%
81	»	93	0.01%
82	«	85	0.01%
83	“	82	0.01%
84	<U+203C>	78	0.01%
85	↷	70	0.01%
86	←	59	0.01%
87	(^ ^)	55	0.005%
88	<U+2669>	49	0.004%
89	⑧	46	0.004%
90	<U+2022>	40	0.003%

91	⑨	28	0.002%
92	[27	0.002%
93]	27	0.002%
94	ˆ	27	0.002%
95	˘	25	0.002%
96	<U+266C>	25	0.002%
97	<U+25B6>	24	0.002%
98	<	22	0.002%
99	>	22	0.002%
100	⑩	22	0.002%
101	▲	20	0.002%
102	☆彡	20	0.002%
103	∇	16	0.001%
104	<U+21E8>	16	0.001%
105	↗	16	0.001%
106	{	14	0.001%
107	}	14	0.001%
108	<U+25B7>	13	0.001%
109	\$	13	0.001%
110	(^ _ ^)	12	0.001%
111	÷	11	0.001%
112	<U+21E9>	11	0.001%
113	∨	10	0.001%
114	⑤	10	0.001%
115	⑪	8	0.001%
116	い	8	0.001%
117	”	6	0.001%
118	<U+25C9>	6	0.001%
119	<U+2666>	6	0.001%
120	±	5	0.0004%
121	<U+2049>	5	0.0004%
122	<U+25E6>	5	0.0004%
123	⑫	5	0.0004%
124	m ²	5	0.0004%
125	ˆ ˆ ;	4	0.0003%
126	<U+2665>	4	0.0003%
127	┆	4	0.0003%
128	〒	4	0.0003%
129	\	4	0.0003%
130	⑬	4	0.0003%
131	-	3	0.0003%
132	(^ ^)	3	0.0003%
133	≡	3	0.0003%
134	≧	3	0.0003%
135	’	2	0.0002%
136	(株)	2	0.0002%
137	°	2	0.0002%
138	↔	2	0.0002%
139	<U+21C4>	2	0.0002%
140	<U+24F5>	2	0.0002%
141	①	2	0.0002%
142	え	2	0.0002%
143	”	1	0.0001%
144	”	1	0.0001%
145	<U+3016>	1	0.0001%

146	<U+3017>	1	0.0001%
147	<U+3018>	1	0.0001%
148	<U+3019>	1	0.0001%
149	<U+FF5F>	1	0.0001%
150	<U+FF60>	1	0.0001%

151	><	1	0.0001%
152	√	1	0.0001%
153	┌	1	0.0001%
154	#	1	0.0001%
155	m	1	0.0001%

資料3 competition 型の発注文書で使用されていた記号の種類と頻度

	語彙素	頻度	割合 (%)
1	。	5808	15.29%
2	、	5774	15.20%
3	・	4419	11.63%
4	/	3267	8.60%
5	.	2298	6.05%
6	:	1959	5.16%
7)	1889	4.97%
8	(1887	4.97%
9	▽	1188	3.13%
10	※	1181	3.11%
11	■	1172	3.08%
12	┌	813	2.14%
13	└	812	2.14%
14	【	626	1.65%
15	】	618	1.63%
16	=	545	1.43%
17	%	543	1.43%
18	!	434	1.14%
19	～	330	0.87%
20	—	240	0.63%
21	×	235	0.62%
22	&	172	0.45%
23	?	158	0.42%
24	●	103	0.27%
25	』	101	0.27%
26	『	100	0.26%
27	+	94	0.25%
28	◆	91	0.24%
29	○	84	0.22%
30	→	81	0.21%
31	▼	78	0.21%
32	—	77	0.20%
33	#	57	0.15%
34	<	54	0.14%
35	>	54	0.14%
36	★	52	0.14%
37	”	50	0.13%
38	*	48	0.13%
39	◎	42	0.11%
40	/	37	0.10%
41]	35	0.09%
42	[32	0.08%
43	...	29	0.08%
44	⇒	24	0.06%
45	cm	23	0.06%

46	～	22	0.06%
47	○	22	0.06%
48	unk	20	0.05%
49	□	19	0.05%
50	¥	16	0.04%
51	◇	15	0.04%
52	《	11	0.03%
53	》	11	0.03%
54	↓	10	0.03%
55	“	9	0.02%
56	←	9	0.02%
57	▲	7	0.02%
58	☆	7	0.02%
59	mm	7	0.02%
60	@	6	0.02%
61	ˆ	6	0.02%
62	,	5	0.01%
63	”	5	0.01%
64	-	4	0.01%
65	;	4	0.01%
66	ˆ	4	0.01%
67	♪	4	0.01%
68	≪	3	0.01%
69	≫	3	0.01%
70	○	3	0.01%
71	〒	3	0.01%
72	m ²	3	0.01%
73	っ	3	0.01%
74	(株)	2	0.01%
75	{	2	0.01%
76	}	2	0.01%
77	↑	2	0.01%
78	<U+25B6>	2	0.01%
79	<U+2777>	2	0.01%
80	④	2	0.01%
81	⑥	2	0.01%
82	—	1	0.003%
83	(株)	1	0.003%
84	<	1	0.003%
85	>	1	0.003%
86	[1	0.003%
87]	1	0.003%
88	\	1	0.003%
89	`	1	0.003%
90	<U+203C>	1	0.003%
91	<U+21E8>	1	0.003%
92	<U+25B7>	1	0.003%

93	<U+25C9>	1	0.003%
94	<U+2661>	1	0.003%
95	<U+2776>	1	0.003%
96	≡	1	0.003%
97	△	1	0.003%

98	☆彡	1	0.003%
99	⑦	1	0.003%
100	⑧	1	0.003%
101	i	1	0.003%

資料4 task型の発注文書で使用されていた記号の種類と頻度

	語彙素	頻度	割合 (%)
1	。	164034	18.89%
2	、	128230	14.77%
3	・	103096	11.87%
4	※	38618	4.45%
5	=	36925	4.25%
6	【	34057	3.92%
7	】	34042	3.92%
8)	32696	3.77%
9	(31561	3.63%
10	/	27378	3.15%
11	.	24875	2.86%
12	」	22155	2.55%
13	「	21937	2.53%
14	:	21826	2.51%
15	■	17737	2.04%
16	～	16433	1.89%
17	!	13135	1.51%
18	—	12611	1.45%
19	?	6501	0.75%
20	*	4717	0.54%
21	—	4586	0.53%
22	—	4583	0.53%
23	○	4192	0.48%
24	%	3768	0.43%
25	▽	3762	0.43%
26	...	3393	0.39%
27	★	3245	0.37%
28	>	3124	0.36%
29	<	3088	0.36%
30	●	2957	0.34%
31	◆	2877	0.33%
32	&	2669	0.31%
33	○	2496	0.29%
34	『	2186	0.25%
35	』	2183	0.25%
36	△	2044	0.24%
37	—	1929	0.22%
38	▼	1889	0.22%
39	☆	1698	0.20%
40	→	1509	0.17%
41	○	1471	0.17%
42	♪	1455	0.17%
43	^	1358	0.16%
44	,	1194	0.14%
45	⇒	1128	0.13%

46	×	1106	0.13%
47	↓	994	0.11%
48	+	841	0.10%
49		710	0.08%
50]	652	0.08%
51	[635	0.07%
52	”	621	0.07%
53	◎	495	0.06%
54	~	478	0.06%
55	<U+2022>	403	0.05%
56	◇	394	0.05%
57	□	257	0.03%
58	《	249	0.03%
59	》	249	0.03%
60	⊥	233	0.03%
61	←	222	0.03%
62	/	210	0.02%
63	↑	205	0.02%
64	▲	194	0.02%
65	(^ ^)	192	0.02%
66	⑥	168	0.02%
67	^ ^	142	0.02%
68	#	117	0.01%
69	@	108	0.01%
70	④	108	0.01%
71	“	90	0.01%
72	unk	88	0.01%
73	⑦	58	0.01%
74	っ	53	0.01%
75	`	46	0.01%
76	´	46	0.01%
77	<<	42	0.01%
78	>>	42	0.01%
79	⑧	42	0.01%
80	ゞ	35	0.004%
81	①	31	0.004%
82	<U+2661>	28	0.003%
83	え	25	0.003%
84	<	24	0.003%
85	>	24	0.003%
86	⑨	23	0.003%
87	<U+2666>	20	0.002%
88	—	18	0.002%
89	;	15	0.002%
90	\	14	0.002%
91	⑩	14	0.002%
92	{	12	0.001%
93	}	12	0.001%

94	∞	9	0.001%
95	②	9	0.001%
96	<U+2665>	8	0.001%
97	<U+3016>	8	0.001%
98	<U+3017>	8	0.001%
99	⑤	8	0.001%
100	∀	7	0.001%
101	⑪	7	0.001%
102	—	6	0.001%
103	-	6	0.001%
104	[6	0.001%
105]	6	0.001%
106	<U+2013>	6	0.001%
107	<U+21E8>	6	0.001%
108	③	6	0.001%
109	<U+2669>	5	0.001%
110	々	4	0.001%
111	い	4	0.001%
112	う	4	0.001%
113	(・∀・)	3	0.000%
114	<U+25B6>	3	0.0003%
115	<U+2776>	3	0.0003%

116	<U+2777>	3	0.0003%
117	<U+2778>	3	0.0003%
118	<U+2779>	3	0.0003%
119	<U+277A>	3	0.0003%
120	‘	2	0.0002%
121	ˆ ˆ ;	2	0.0002%
122	<U+216A>	2	0.0002%
123	><	2	0.0002%
124	≠	2	0.0002%
125	∴	1	0.0001%
126	’	1	0.0001%
127	ˆ ∇ ˆ	1	0.0001%
128	°	1	0.0001%
129	±	1	0.0001%
130	<U+203C>	1	0.0001%
131	><)	1	0.0001%
132	¯	1	0.0001%
133	∨	1	0.0001%
134	⑫	1	0.0001%