程度を考慮したフォーマリティ変換のための データセットの収集と分析

守屋彰二 [1]　岸波洋介 [1]　佐藤志貴 [1]　徳久良子 [1]　乾健太郎 [1,2]
[1] 東北大学 [2] 理化学研究所
{shoji.moriya.q7,yosuke.kishinami.q8}@dc.tohoku.ac.jp
{shiki.sato.d1,tokuhisa,kentaro.inui}@tohoku.ac.jp

概要

自然言語処理におけるテキストスタイル変換では，文の意味を保持したまま両極端の 2 つのスタイルの一方から他方へ変換することが一般的である．それに対して我々は，連続的なスタイルの程度を制御可能な変換の実現を目指す．本研究ではその第一歩として，フォーマリティの程度が異なる同内容の文に対し，その度合いが付与されているデータセット『フォーマリティスコア付き GYAFC』を構築した．分析の結果，全体としてフォーマリティの程度について人の間で一定の合意が得られること，誤記やスラングによる程度の差異に関しては特に合意が得やすい可能性があることがわかった．本研究で収集したデータセットは，後日公開予定である．

1　はじめに

近年，自然言語処理の分野において，テキストの意味を保ちながらフォーマリティやシンプリシティなどのスタイルを変換するテキストスタイル変換の研究が盛んに行われている [1, 2]．テキストスタイル変換は，状況に応じたテキストを生成する上で非常に有用な技術であり，幅広い用途があることで知られている [3, 4]．例えば，学習データの水増し [5] や一貫したペルソナをもつチャットボットへの応用 [6] などがある．

現在行われているテキストスタイル変換の研究の多くは，フォーマル・インフォーマルのように両極端のスタイルの一方からもう一方へと変換するものである [1]．しかし，現実では「**スタイルの程度**」が存在すると考えられる．例えば，相手との距離感が近くなるにつれ，用いる表現のフォーマリティの程度が徐々に低くなっていくことが想定される．程度を考慮した上でスタイル変換することは実応用における融通性・有用性を高める上で非常に重要な観点であると考えられる．

表 1　構築した『フォーマリティスコア付き GYAFC』

スコア	文
6.4	I think this is the best way
5.2	I think that’s the way to go
2.4	best way i think
1.6	i this this way is good.
6.3	Because I feel so lonely in the school
5.3	Because feel more lonely in school
4.0	Because i feel alone in the school
2.0	Coz i feel so lonely in the school..

しかし，そもそも「スタイル」には一般的な定義は存在せず，個人の直観に基づく概念であるとされている [3]．そのため，連続的に（完全に連続でなくてもスケーラブルに）変化すると考えられる「スタイルの程度」に関しても，人の間で合意が取れるかどうかは自明ではない．仮に人の間でまったく合意が取れなければ，スタイルの程度を変化させた文を自動生成することは難しい．逆に，ある程度人の間で合意が取れれば，人手によりコーパスを整備することで，連続的に程度を制御可能なスタイル変換システムを構築できると考えられる．

そこで，本研究では，スタイルの一つであるフォーマリティに着目し，フォーマリティの程度に対する共通認識を形成できるかについて調べる．具体的には，まず，フォーマリティの程度が異なる同じ意味の文に対してその度合いが付与されたデータを小規模に収集した (表 1)．その後，フォーマリティスコアをもとにワーカ間でフォーマリティの程度に関する合意が取れるかどうかを調べる．

本研究の貢献は，以下の通りである．1) 多様な程度のフォーマリティで書かれた同じ意味の文に対し，その程度が付与されたデータの収集手法を考案

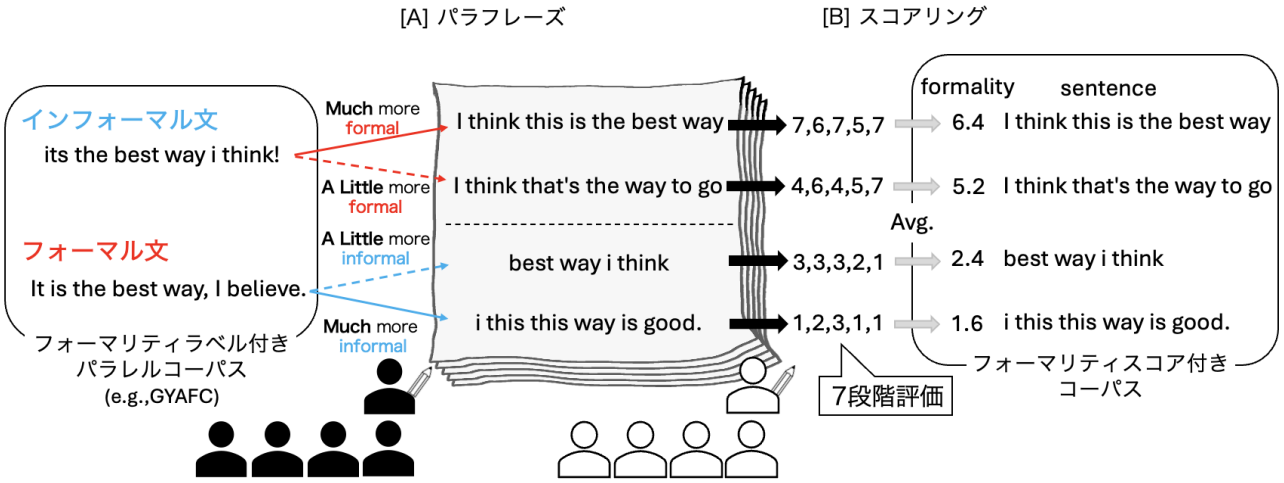


図 1　提案手法：フォーマリティスコア付きコーパスの収集

する．2) フォーマリティの程度が付与されている 1100 文のデータセット『フォーマリティスコア付き GYAFC』を構築する．3) フォーマリティの程度に対して，人の間で一定の認識の合意が取れるかどうかを確認し，合意の得やすさに影響を与える要素について調べる．

2　関連研究

2.1　テキストスタイルの定義

テキストスタイル変換の研究におけるスタイルは，個人が持つ固有の特性を表現する方法という直観的な概念であるとされている [3]．今回我々が扱うフォーマリティについても，一般的な定義は存在しない [7]．社会的な距離や共有された知識などの状況によって定義される例もあれば [4]，スラングの使用や文法的な誤りなどのより抽象度の低い定義が採用される例もある [8, 9]．

2.2　フォーマリティのデータセット

フォーマリティのラベル付きパラレルコーパスとして，フォーマル・インフォーマルのスタイルラベルが付与されている GYAFC (Google Yahoo Answer Formality Corpus) [10] や，複数ヶ国の言語が含まれる XFORMAL [11] などがある．また，フォーマリティの程度がアノテーションされているノンパラレルコーパスとして，formality-corpus [7] がある．

フォーマリティの程度に対する合意を調べる上で，異なる意味の文を比較するよりも同じ意味の文を比較する方が，相対的な程度の差が明確になると考えられる．そこで本研究では，同じ意味の複数の文に対しフォーマリティの程度が付与されたデータセット『フォーマリティスコア付き GYAFC』を構築した．

3　提案手法

同じ意味の複数の文に対しフォーマリティの程度が付与されたデータの収集方法を提案する．まず，多様な程度のフォーマリティを持つ文を収集するため，フォーマリティのラベルのついたパラレルコーパスの文に対しパラフレーズを行う．その後，収集した文にフォーマリティスコアを付与する．なお，本手法では Pavlick ら [7] に従い，フォーマリティとは何であるかを具体的に明記せず，各個人が持っている独自のフォーマリティの定義に従うボトムアップの手法を採用した．

パラフレーズ（図 1 の [A]）．　まず，GYAFC のうちフォーマル文をワーカに与え，その文と同じ意味の「A Little more informal」「Much more informal」の 2 種類のインフォーマル文を書かせる．これを各文あたり 5 人のワーカに行わせ，合計 10 文のインフォーマル文を得る．同様に，インフォーマル文に対してもワーカに「A Little more formal」「Much more formal」の 2 種類のフォーマル文を書かせ，合計 10 文のフォーマル文を得る．以上より，コーパスに元々収録されているフォーマル・インフォーマルの 2 文に加えて 20 文を得る．以下，この 22 文からなる集合を 1 グループとする．2.1 節で述べた通り，フォーマリティは各個人が持っている独自の定義に従うことから，複数のワーカに同じ設定でパラフレーズを行わせることで多様な程度のフォーマリティを持つ文が収集可能と考えられる．

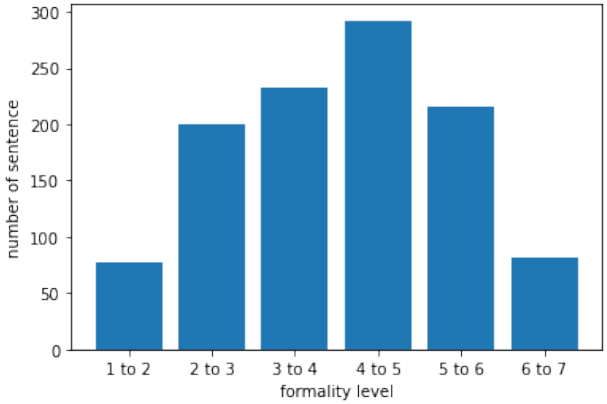


図 2　収集した文のフォーマリティスコアの分布

スコアリング（図 1 の [B]）．　各グループに含まれる文に対し，Pavlick ら [7] と同様にワーカに以下の 7 段階でフォーマリティスコアを付与する:

- 1-Very Informal
- 2-Moderately Informal
- 3-A Little Informal
- 4-Neutral
- 5-A Little Formal
- 6-Moderately Formal
- 7-Very Formal

このスコアリングを 1 グループあたり 5 人のワーカに行わせ，文に付与された 5 人のフォーマリティスコアの平均値を最終的な文のフォーマリティスコアとする．このようにして，多様な程度のフォーマリティで書かれた同じ意味の文に対し，フォーマリティスコアを付与したデータを収集する．なお，2.1 節で述べたように各個人が持っている独自のフォーマリティの定義に従うため，ワーカへの教示は必要最低限にとどめる．[1)]

今回はフォーマリティに着目してコーパスを構築したが，本収集手法はフォーマリティに限らず他の種類のスタイルにも適用できる可能性があると考えられる．具体的には，「シンプリシティ」や「ポライトネス」のように，連続的に変化するスタイルについては今回提案した収集方法が適用できる可能性がある．

4　データ収集の結果と分析

提案手法をもとに小規模に収集を行い，『フォーマリティスコア付き GYAFC』を構築した．収集したデータの同グループ内の文のスコアを比較するこ

1）　本研究でデータセットを構築した際の教示画面を付録 A に示す．

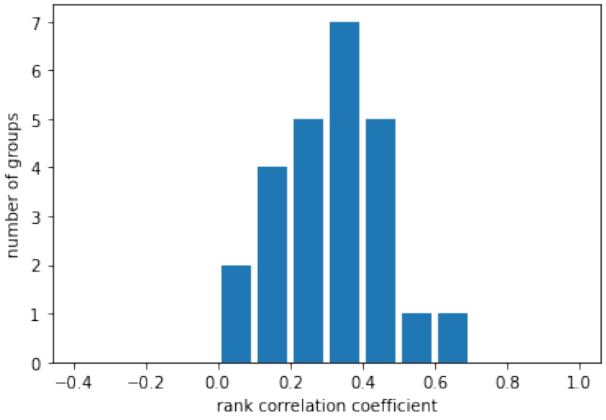


図 3　グループごとの順位相関係数の分布

とで，人の間で相対的なフォーマリティの程度について合意が取れるかどうかを調べた．また，合意の得やすさに影響を与える要素について調べた．

4.1　データ収集の設定

パラフレーズ．　GYAFC [10] から無作為に抽出したフォーマル文とインフォーマル文の 50 対に対し，パラフレーズを行った．収集には Amazon Mechanical Turk[2)] を使用した．[3)]

スコアリング．　GYAFC [10] の 50 対から作成した全てのグループに含まれる文に対し，フォーマリティスコアの付与を行った．収集には，Amazon Mechanical Turk[2)] を使用した．[4)]

ワーカのフィルタリング．　アノテーションの質を担保するため，MACE[5)] を用い，competence の値が低い下位 30%のワーカを除去した上で，平均値を計算して各文のフォーマリティスコアを得た．

4.2　収集結果

多様な程度のフォーマリティの文を計 1100 文収集し，『フォーマリティスコア付き GYAFC』を構築した．収集した文のフォーマリティスコアの分布を図 2 に示す．収集した文のフォーマリティスコアが幅広く分布していることから，さまざまなフォーマリティの程度を持つ文が収集できていることが確認できた．また，実際に収集したデータの文例を表 1 に示す．

2）　https://www.mturk.com/

3）　97%以上の承認率，1000 件以上のタスク承認歴をもつワーカを 1HIT あたり$3.5 で 65 人雇用した．1HIT あたり 4 文対のパラフレーズを行わせた．想定時間は約 15 分であった．

4）　パラフレーズと同様の条件でワーカを 1HIT あたり$3.5 で 250 人雇用した．1HIT あたり 1 グループ (22 文) スコアリングを行わせた．想定時間は約 15 分であった．

5）　https://github.com/dirkhovy/MACE

表 2　合意の取りやすさに影響を与える要因 (要因が含まれる箇所を太字で記す)

グループ	スコア	文	順位相関係数
(a)	6.4	I think this is the best way	0.450
	3.6	**its** the best way **i** think!	
	1.6	**i this** this way is good.	
(b)	4.8	Really it is Sharal crow	0.381
	3.8	**Ya** of corse that sharal crow sound	
	1.0	**yaaaaaaaaaaaaaaaaaa** of corse that sharal crow **girrrrrrrrrrrrrl**	
(c)	6.0	I think Vin Diesel is **really attractive**	0.265
	4.3	I see Vin Diesel is **more sexy**.	
	3.8	I think Vin Diesel may be **so hot**	

4.3　分析

同じグループ内に含まれる文のフォーマリティスコアを比較することで，相対的なフォーマリティの程度について人の評価がどの程度一致するかを調べる．また，フォーマリティの程度に対して合意が取れるグループにはどのような特徴があるかを分析する．

フォーマリティの程度に対する合意．　人の間で相対的なフォーマリティの程度に合意が取れるかを調べるため，同じ文に対しスコアリングを行ったワーカ間でフォーマリティの順位相関係数を算出した．なお，今回は各グループごとに 5 人の異なるアノテータが存在するため，5 人の全組み合わせにおける，各ワーカのフォーマリティスコアに基づくグループ中の文のランキングの順位相関係数を計算した[6]．各グループの順位相関係数の分布を図 3 に示す．平均値は 0.321，中央値は 0.333 であり，ワーカの順位の間で正の相関が見られた．この結果から，フォーマリティの程度についてワーカ間で一定の合意が得られることが確認できた．一方で，グループ間で順位相関係数にばらつきが見られ，グループによって合意の取りやすさが異なることも確認できた．

合意の取りやすさに影響を与える要因．　収集したデータを定性的に確認し，合意の取れやすさに影響を与える言語的要因について分析した．順位相関係数が高いグループでは，表 2 の (a) や (b) のように誤記やスラングの有無や多さによるフォーマリティの程度の差がみられた．一方で，順位相関係数が低いグループでは，表 2 の (c) のように句の言い換えによってフォーマリティの程度に差が生じる例がみられた．このような分析は，今回我々が作成した，フォーマリティが異なる同じ意味の文からなるコーパスを用いることで可能になるものと考える．

6)　ワーカのフィルタリングにより，1 グループあたりのアノテータが 3 人以下となったグループは除外した．

5　おわりに

程度を考慮したスタイル変換を行うことは，実応用において重要であると考えられる．しかし，「スタイルの程度」に関して，人の間で合意が取れるかどうかは自明ではない．本研究では，スタイルの一つであるフォーマリティに着目し，フォーマリティの程度に対して合意が取れるかどうか調べた．具体的には，フォーマリティの程度が異なる同じ意味の文に対してその度合いが付与されたデータを小規模に収集し，分析を行った．その結果，フォーマリティの程度について一定の認識の合意が得られることが確認できた．また，誤記やスラングによる程度の際に関しては，特に合意が得やすい可能性があることがわかった．

今回の収集では，7 段階のスコアのランキングをもとに順位相関係数を計算し，フォーマリティの程度に対する認識の一致を確認した．しかし，ワーカ間で絶対的なスコアの基準を合わせることは難しく，そのスコア自体の信頼性は十分とは言えない．今後の研究で，より適したアノテーションの方式について模索していきたい．また，本研究で提案した手法は，フォーマリティに限らず他の種類のスタイルにも適用できる可能性があると考えている．他の種類のスタイルにおいてもフォーマリティと同様に程度の概念が存在しているのかどうかや，人々の間でその程度に対して合意が得られるかについて調査していきたい．

謝辞

本研究は，JSPS 科研費 JP21J22383，JST CREST JPMJCR20D2 の助成を受けたものです．本研究を進めるにあたり，愛媛大学の梶原智之先生に有益なご助言をいただきました．また，AMT のタスク作成にあたっては，東北大学乾研究室の藤原吏生氏，Steven Coyne 氏，栗田宙人氏に有益なコメントをいただきました．ここに深く感謝申し上げます．

参考文献

[1] Di Jin, Zhijing Jin, Zhiting Hu, Olga Vechtomova, and Rada Mihalcea. Deep Learning for Text Style Transfer: A Survey. **Computational Linguistics**, Vol. 48, No. 1, pp. 155–205, 04 2022.

[2] Tomoyuki Kajiwara, Biwa Miura, and Yuki Arase. Monolingual transfer learning via bilingual translators for style-sensitive paraphrase generation. **Proceedings of the AAAI Conference on Artificial Intelligence**, Vol. 34, pp. 8042–8049, 04 2020.

[3] David D. McDonald and James D. Pustejovsky. A computational theory of prose style for natural language generation. In **Proceedings of the Second Conference on European Chapter of the Association for Computational Linguistics**, EACL '85, p. 187–193, USA, 1985. Association for Computational Linguistics.

[4] Eduard Hovy. Generating natural language under pragmatic constraints. **Journal of Pragmatics**, Vol. 11, No. 6, pp. 689–719, 1987.

[5] Yi Zhang, Tao Ge, and Xu Sun. Parallel data augmentation for formality style transfer. In **Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics**, pp. 3221–3228, Online, July 2020. Association for Computational Linguistics.

[6] Yanpeng Zhao, Wei Bi, Deng Cai, Xiaojiang Liu, Kewei Tu, and Shuming Shi. Language style transfer from sentences with arbitrary unknown styles. **CoRR**, Vol. abs/1808.04071, , 2018.

[7] Ellie Pavlick and Joel Tetreault. An empirical analysis of formality in online communication. **Transactions of the Association for Computational Linguistics**, Vol. 4, pp. 61–74, 2016.

[8] Alejandro Mosquera and Paloma Moreda. A qualitative analysis of informality levels in web 2.0 texts: The facebook case study. pp. 23–29, 01 2012.

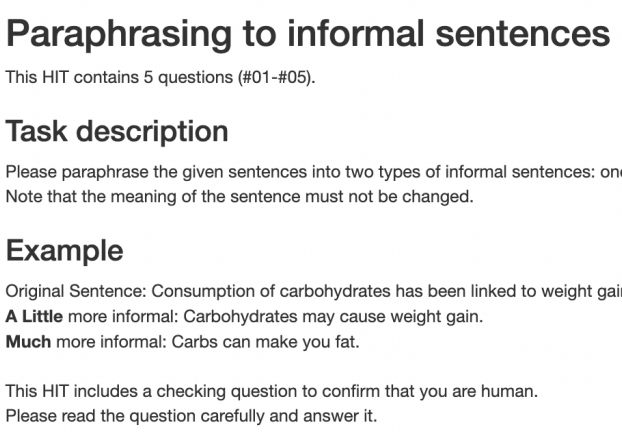
[9] Kelly Peterson, Matt Hohensee, and Fei Xia. Email formality in the workplace: A case study on the Enron corpus. In **Proceedings of the Workshop on Language in Social Media (LSM 2011)**, pp. 86–95, Portland, Oregon, June 2011. Association for Computational Linguistics.

[10] Sudha Rao and Joel Tetreault. Dear sir or madam, may I introduce the GYAFC dataset: Corpus, benchmarks and metrics for formality style transfer. In **Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)**, pp. 129–140, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.

[11] Eleftheria Briakou, Di Lu, Ke Zhang, and Joel Tetreault. Olá, bonjour, salve! XFORMAL: A benchmark for multilingual formality style transfer. In **Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies**, pp. 3199–3216, Online, June 2021. Association for Computational Linguistics.

A　クラウドワーカへの教示画面

4.1 節で述べたパラフレーズおよびスコアリングにおいてワーカに提示した教示画面を，図 4 および図 5 にそれぞれ示す．



Paraphrasing to informal sentences

This HIT contains 5 questions (#01-#05).

Task description

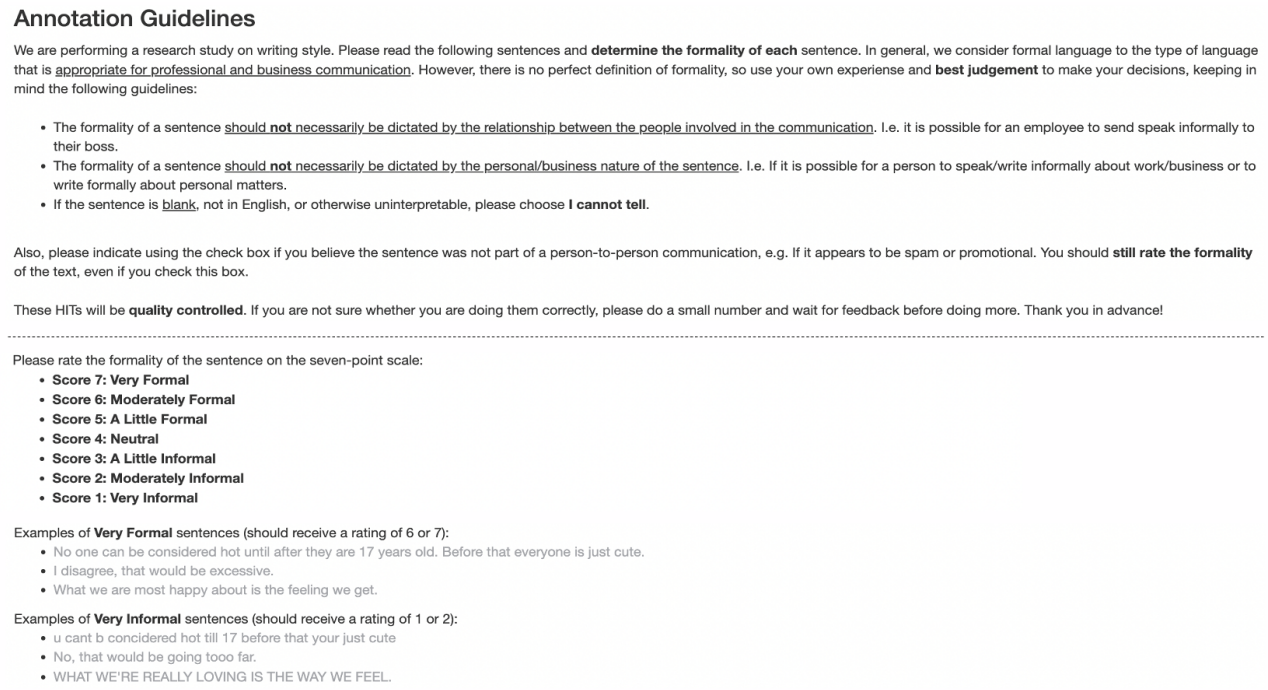
Please paraphrase the given sentences into two types of informal sentences: one is **A Little** more informal, and the other is **Much** more informal.
Note that the meaning of the sentence must not be changed.

Example

Original Sentence: Consumption of carbohydrates has been linked to weight gain.
A Little more informal: Carbohydrates may cause weight gain.
Much more informal: Carbs can make you fat.

This HIT includes a checking question to confirm that you are human.
Please read the question carefully and answer it.

図 4　パラフレーズでのクラウドワーカへの教示画面



Annotation Guidelines

We are performing a research study on writing style. Please read the following sentences and **determine the formality of each** sentence. In general, we consider formal language to the type of language that is appropriate for professional and business communication. However, there is no perfect definition of formality, so use your own experiense and **best judgement** to make your decisions, keeping in mind the following guidelines:

- The formality of a sentence should **not** necessarily be dictated by the relationship between the people involved in the communication. I.e. it is possible for an employee to send speak informally to their boss.
- The formality of a sentence should **not** necessarily be dictated by the personal/business nature of the sentence. I.e. If it is possible for a person to speak/write informally about work/business or to write formally about personal matters.
- If the sentence is blank, not in English, or otherwise uninterpretable, please choose **I cannot tell**.

Also, please indicate using the check box if you believe the sentence was not part of a person-to-person communication, e.g. If it appears to be spam or promotional. You should **still rate the formality** of the text, even if you check this box.

These HITs will be **quality controlled**. If you are not sure whether you are doing them correctly, please do a small number and wait for feedback before doing more. Thank you in advance!

Please rate the formality of the sentence on the seven-point scale:

- **Score 7: Very Formal**
- **Score 6: Moderately Formal**
- **Score 5: A Little Formal**
- **Score 4: Neutral**
- **Score 3: A Little Informal**
- **Score 2: Moderately Informal**
- **Score 1: Very Informal**

Examples of **Very Formal** sentences (should receive a rating of 6 or 7):

- No one can be considered hot until after they are 17 years old. Before that everyone is just cute.
- I disagree, that would be excessive.
- What we are most happy about is the feeling we get.

Examples of **Very Informal** sentences (should receive a rating of 1 or 2):

- u cant b concidered hot till 17 before that your just cute
- No, that would be going tooo far.
- WHAT WE'RE REALLY LOVING IS THE WAY WE FEEL.

図 5　スコアリングでのクラウドワーカへの教示画面