

ゲーミフィケーションによる連想概念の獲得

Acquiring word associations with Gamification

町田雄一郎[†]河原大輔[†]柴田知秀[†]黒橋禎夫[†]颯々野学[‡]

Yuichiro Machida

Daisuke Kawahara

Tomohide Shibata

Sadao Kurohashi

Manabu Sassano

1 研究背景

近年、自然言語処理では大規模なリソースを利用した手法が大きな成果を上げており、充実した言語リソースやデータセットの整備は自然言語処理研究においては欠かせない。このようなリソースの構築には主に2つのアプローチが取られる。1つは人手で構築する手法で、もう1つは自動的に構築する手法である。人手で作成されたりリソースとしては WordNet[4] や FrameNet[1] などがある。これらは専門家によって作成されているため一般的に質は高いが、大規模なリソースの作成には多くの時間とコストがかかってしまうことが問題である。一方、自動的に構築されたりリソースとしては、日本語格フレーム [14] などがある。自動的な手法を用いる場合、大規模なリソースを構築できる反面、人手で作成するリソースよりも質が低いことが問題である。また、大規模であるがゆえに全ての結果を人手でチェックすることも難しい。

近年では、インターネットサービスを介して世界中の不特定多数の人間に作業を依頼することのできるクラウドソーシングを利用したリソースの構築も盛んである。クラウドソーシングを利用することで、大規模な作業でも素早く、かつ低コストで行うことが可能である。しかしながら、一般的にクラウドソーシングで複雑なタスクを依頼する際には、成果物の質を高めるために、本来行いたい作業を細かく簡単な作業に分割することが推奨されている。そのため、複雑なアノテーションを大規模に行う際には、実際の作業規模が増大してしまうという問題がある。従って作業員1人あたりの単価は低くとも、全体的なコストは依然として問題となる場合がある。

一方、ゲームで遊ぶ行為そのものが、何か別の作業となっている仕組みであるゲーミフィケーションを利用する

試みも近年様々な分野で行われている。ゲーミフィケーションを効果的に用いることができれば、作業のコストを更に抑えることが可能になる。

自然言語処理では様々なリソースが利用されているが、その一つとして語と語の関係知識に着目して作成されたりリソースがあげられる。特に、ある語に関する連想概念を獲得することは、言葉の背景にある情報を扱うことができるため、より高度な自然言語処理技術の発展に貢献することが期待される。

そこで本研究では連想概念を対象に、自動的な構築手法、ゲーミフィケーション、クラウドソーシングを効果的に組み合わせ、質が高く、大規模な言語リソースを安価に構築するための手法を提案する。具体的には、大規模 Web コーパスを用いて自動獲得した連想概念に対して、対話型エージェントとの連想ゲームという形式で、獲得結果の質を判定する。質が悪いと判定された連想概念に対しては、クラウドソーシングを用いて不特定多数の人間に正しい結果に修正させる。

2 関連研究

2.1 言語処理におけるクラウドソーシング

Snow ら [8] は様々な自然言語処理のアノテーションタスクをクラウドソーシングで行い、専門家によるものと比較したところ、手法を工夫すれば高い精度のアノテーションが獲得できることを示した。これを発端に多くの自然言語処理タスクをクラウドソーシングで行うという研究がなされている。例えば、推論規則の評価 [13]、機械翻訳の評価 [2]、クラウドソーシングを利用して高い質の翻訳を獲得する [12] など、多くの研究が行われている。更に評価用のデータセットの作成などにも盛んにクラウドソーシングが利用されてきている。大規模な言語リソースをクラウドソーシングを用いて作成する試みも行われており、FrameNet をクラウドソーシングを用いて構築する研究 [5] や、談話関係の判定をクラウドソーシングを利用して判定する研究 [6] も行われている。

[†] 京都大学大学院情報学研究所, Graduate School of Informatics, Kyoto University

[‡] ヤフー株式会社, Yahoo Japan Corporation

2.2 ゲーミフィケーション

クラウドソーシングにおいて、ワーカーが作業を行う動機は様々である。例えば Amazon Mechanical Turk などのクラウドソーシングサービスに参加するユーザーは主に報酬を得ることを目的としている。また Wikipedia などのサービスもクラウドソーシングの一種であると考えられているが、記事の編集を行うユーザーはサービスへの貢献などを動機として編集を行っている。ゲーミフィケーションは、ユーザーが楽しみを享受することを目的として参加するクラウドソーシングの一形態である。ゲーミフィケーションの仕組みは von Ahn[11] によって提案され、Game with a Purpose という言葉で表現されている。ゲーミフィケーションでは、プレイヤーがゲームを楽しむその行為自体が、何か別の作業となっているという特徴がある。またユーザーは楽しみを享受することを目的としてゲームを行うため、ゲーミフィケーションは基本的には無料で行うことができる。従ってゲーミフィケーションを効果的に用いることができれば、作業にかかるコストを大幅に減らすことが可能である。Fold.it[3] ではタンパク質の構造解析をゲームとしてユーザーに行わせた。参加したユーザーは数十万人規模で、研究者が長い間解くことができなかったタンパク質の構造解析に成功するという成果を上げており、ゲーミフィケーションを効果的に利用することは大きな成果に繋がることが示されている。自然言語処理の分野においては、ESP ゲームを行うことで Semantic Network を作成することや [7]、語義のラベル付けを行う [10] などといった研究がされている。最近では、関連語のアノテーションをビデオゲームの形で行うといった研究 [9] も行われている。この研究では、報酬型クラウドソーシングでアノテーションを行った時よりもゲーミフィケーションを利用した方が精度が向上したことが報告されている。

3 提案手法

本研究の目的は、自動構築、ゲーミフィケーション、クラウドソーシングを効果的に組み合わせることで、大規模かつ、質の高い連想概念リソースを低コストで作成することである。具体的には、まず進らの手法 [15] を用いて連想概念のリソースを自動構築する。ここで本論文における連想概念とは、多くの人間がある名詞から連想すると思われる単語集合を指す。進らはある名詞に対して関連する名詞集合を自動獲得する手法を提案しているが、本研究では更に、ある名詞とそれに関連する形容詞も連想概念とし扱う。しかし、このようにして自動的に獲得された連想

概念の中には正しくないものも含まれている。このような質の悪い連想概念を音声対話と連想ゲームを利用したゲーミフィケーションでチェックする。大規模に自動獲得された連想概念を修正するには、どの獲得結果の質が悪いかを把握した上で、別の連想概念に修正しなければならない。一つの方法としてクラウドソーシングを利用してこれらの作業を行うことが考えられるが、そのためには一連の修正の流れを誰でもできるような小さな作業に分割し、更に回答の信頼性を担保するために複数のワーカーからの判定を得る必要がある。このような形で大規模に修正作業を行うと、かなりの人的資源が必要とされる上、安価なクラウドソーシングといえどもコストがかさんでしまう。しかし作業を分割し、その一部をゲームとしてプレイヤーに提供することでコストを大幅に下げることが可能である。特に、自動獲得の結果を判定する作業は、人間にとっては簡単ではあるが数が多い。この作業をゲーミフィケーションで無料で行うことができれば、コストを大きく下げることができる。このゲームの結果として得られた質の悪い連想概念については、報酬支払型のクラウドソーシングサービスを利用して、不特定多数のユーザーに正しい連想概念を記述してもらう。

3.1 連想概念の自動獲得

連想概念の自動獲得の手法としては進ら [15] のものを用いる。進らは Web テキストと、Wikipedia や辞書の語釈文という 2 種類の知識源を用いて名詞の連想概念を獲得している。まず、一般のテキストにおいて一定の文脈内で共起しやすい語のペアは強い関係を持つことが多いことから、Web テキストを知識源として用いて名詞の共起頻度を計算する。しかし、Web テキストにおける共起の関係だけでは低頻度語に関して関連語を獲得するのが難しいため、名詞の説明をしている文章も知識源として利用する。このような説明箇所として、国語辞典の語釈文や Wikipedia 記事の第一段落などを利用している。また、Wikipedia の曖昧性回避ページを利用して、多義性の判定も行っており、「WBC」のように、世界ボクシング評議会 (World Boxing Council) とワールド・ベースボール・クラシック (World Baseball Classic) のような違いも考慮している。進らはこの手法で 100 万語以上の名詞を対象とした連想概念を獲得している。進らの手法では連想概念として名詞集合を獲得していたが、本研究では、名詞関連語に加えて名詞に関連する形容詞も対象として連想概念を自動獲得する。形容詞の獲得手法としては名詞との共起から自己相互情報量を計算することで獲得する。例として、「ハム」に対して獲得できた形容詞を表 1 に示す。

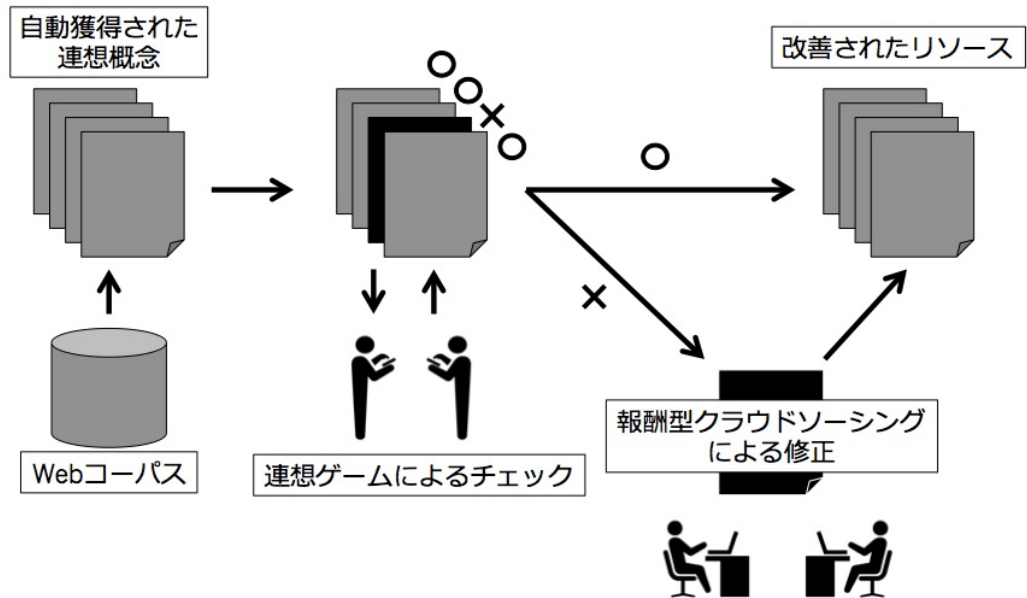


図1 提案手法全体像

表1 「ハム」に対して自動獲得された形容詞の関連語の例

自動獲得された形容詞	相互情報量
美味だ	5.998
美味しい	5.190
やばい	4.105
旨い	4.019
有名だ	3.875

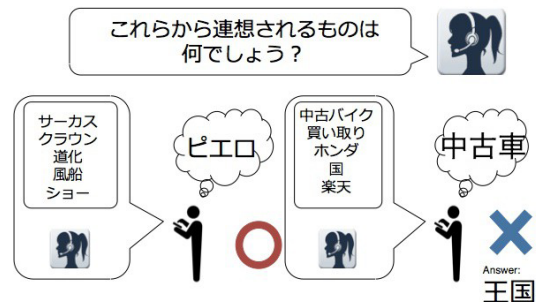


図2 連想ゲームの例

3.2 音声対話を利用した連想ゲーム

自動的に獲得された連想概念の質を連想ゲームとしてプレイヤーに評価させる。プレイヤーとは連想ゲームで遊ぶ人を指す。例えばキーワード「ピエロ」に関する連想概念の獲得結果が、「サーカス・クラウン・道化・風船・ショー」であったとき、プレイヤーにはヒントとして連想概念の「サーカス・クラウン・道化・風船・ショー」が提示され、キーワードが何かを連想してもらおう。ここでキーワードとは連想概念を獲得する対象の単語を指し、ヒントとは、連想ゲームでプレイヤーに提示する単語集合を指す。一方、「中古バイク・買い取り・ホンダ・国」というヒントが提示された際にプレイヤーが「バイク」という回答をしたとする。キーワードは「王国」であり、プレイヤーは期待している単語とは大きく異なった回答をしたことになる。このような場合には連想概念の自動獲得の室が悪かったと判断できる。もし、ある単語の連想概念が正しく獲得できているのであれば、人間はヒントから、キーワードに近い語を推測することができるはずである。

仮に自動獲得の結果の質が悪ければ、人間はキーワードを答えることができないか、または全く異なる語を答えると考えられる。

本研究では、このゲームを Yahoo!が提供している音声対話スマートフォンアプリである「Yahoo!音声アシスト」の追加機能として試験的に動作させることを検討している。^{*1} 追加機能としての導入が実現すれば、ゲームログを収集、分析することでゲーミフィケーションを利用した連想概念の評価実験を行うことが可能となる。また「音声アシスト」を利用することは、話しかけるだけで簡単に遊ぶことができ、ゲームを始めるまでの敷居が低い点や、対話型エージェントとの会話がユーザーの興味を引くという点でも、ゲームを促進する要素を備えていると考えら

*1 <http://v-assist.yahoo.co.jp/>

れる。

3.2.1 連想ゲームの出題方法

自動獲得した連想概念を用いて連想ゲームを行うとき、プレイヤーがゲームをストレスなく行うことができるかは重要な問題となる。例えば提示したヒントの中に、キーワードと非常に類似した単語ばかりが入っていたり、キーワードが抽象的で非常に難解な問題になってしまうと、プレイヤーの意欲を削いでしまう可能性がある。そのため、いくつかの点に留意して出題を行った。まずキーワードのフィルタリングとして次の方法を用いた。Webの頻度で上位から3000~7000位にあるものに限定し、更に形態素解析器JUMANのカテゴリで「抽象物」「時間」が付与されている語を削除した。また、進らの手法では、連想概念に「関連度」というスコアが与えられており、出題は基本的にスコアが大きな順に提示する。同時に、分布類似度を用いてヒント間の類似度を計算し、ヒントとなる単語が互いに似たものにならないように留意した。

3.3 クラウドソーシングを利用した連想概念の修正

連想ゲームの結果から連想概念の自動獲得結果の質が悪く考えられるものについては、正しい連想概念に修正したい。このような場合は報酬支払型のクラウドソーシングを利用することによって不特定多数のユーザーに連想概念を修正してもらう。今回はYahoo!クラウドソーシングを利用した。Yahoo!クラウドソーシングでは不特定多数の日本人のワーカーに、小さな作業を報酬を支払うことで行ってもらうサービスである。ワーカーは1タスクあたり、2~5Tポイント*2を獲得することができ、作業を依頼する側も安価に済ませることが可能である。ここでワーカーとは、インターネットを通して作業を行う作業者を指し、タスクとはワーカーが報酬をもらうためにこなす作業単位を指す。連想概念の修正には2つのステップをとる。第一段階では、ある単語とその連想概念の自動獲得結果が提示され、その中で連想関係にない単語を複数のワーカーにチェックしてもらう。第二段階では、チェックのつかなかった単語、即ち正しく獲得できている単語とは別の連想概念を、ワーカーに記述してもらうという内容である。4節で、このタスクの詳細について説明する。

4 関連語の修正についての予備実験

本稿の段階では、ゲーミフィケーションに関するログデータが集まっていないため、ここでは、ゲーミフィケーションによるログデータが十分集まったことを想定した

上で、次のステップである「修正」と「獲得」についての予備実験の結果を報告する。実験はYahoo!クラウドソーシングを利用して行い、次の2段階のタスクを行った。

連想概念のチェックタスク ある単語の連想概念として自動獲得されたものの内、連想関係にないものを判定する。

連想概念の記述タスク あるキーワードに対する連想概念を自由に記述させる。

このようにタスクを2種類に分けることで、作業工程数が増えてしまうが、クラウドソーシングにおいては作業を細かく分割する方が精度が良い結果が得られるとされているため2段階のステップに分割する必要があると判断した。さらに、どのような単語が誤って獲得されているのかを知ることができるため、自動獲得手法のエラー分析にも貢献でき、今後の研究に有用な情報も同時に獲得できると考えられる。また、「連想概念の記述タスク」については、本来は「連想概念のチェックタスク」の結果を踏まえて行うタスクである。しかし今回はゲームログが無いため、「あるキーワードに対する連想概念を自由に記述させる」という、より単純化した作業内容として実験した。

4.1 連想概念のチェックタスク

4.1.1 連想概念のチェックタスク:実験設定

あるキーワードとそれに対する自動獲得された連想概念をワーカーに提示する。ワーカーは提示された連想概念の内、キーワードからは連想されないものにチェックを付ける。ワーカーにはキーワード1語と、連想概念が10語程度提示される。全体ではキーワードが100語、それに対応する連想概念が各約30語程度の規模でデータセットを構築した。出題例を図3に示す。キーワード1語に対しては、10人のワーカーの判定を獲得する。1タスクは連想概念のチェック作業2題と、ワーカーの信頼性を測るためのチェック問題1題の計3題で構成される。チェック問題はワーカーが人間であることを判定するためのもので、「この設問は1を選択してください」といった設問である。

4.1.2 連想概念のチェックタスク:実験結果

連想概念のチェックタスクに参加したワーカーはのべ217人で、平均実行タスク数は4.47であった。作業時間は約30分であった。チェック問題については99.4%の精度であり、ほとんど全てのワーカーが真剣に取り組んでいると判断した。連想概念のチェックは10人中4人以上のワーカーが関係ないと判断した概念を除外することとした。その結果、自動獲得した連想概念の内、除外された連

*2 <http://tsite.jp/>

元の単語からは“連想されない”単語全てにチェックをつけてください。

元の単語	関節
<input type="checkbox"/> リウマチ	
<input type="checkbox"/> 指	
<input type="checkbox"/> グルコサミン	
<input type="checkbox"/> 腰	
<input type="checkbox"/> コンドロイチン	
<input type="checkbox"/> 可動	
<input type="checkbox"/> 筋肉	
<input type="checkbox"/> 肩	
<input type="checkbox"/> ゴツい	
<input type="checkbox"/> 肘	
<input type="checkbox"/> 痛い	

図 3 連想概念のチェックタスクの例

想概念が 3 割以下であったものは、100 語中 55 語であった。ワーカーによる選別例を表 2 に示す。ここで、下線部はワーカーによる選別の結果、連想概念であると判定されたものである。また判定された結果を著者が目視で検討した結果、ワーカーの判定の集約結果は正しいものであることが確認できた。

4.2 連想概念の記述タスク

4.2.1 連想概念の記述タスク:実験設定

ある単語をワーカーに提示し、その単語から連想される概念をなるべくたくさん記述させる。対象とした単語は 100 語であり、1 単語につき、各 10 人のワーカーに記述させる。また、1 タスクは 1 題の記述作業と 1 題のチェック問題から構成され、1 人のワーカーは最大で 5 回作業をすることができる。設問の形式は次のとおりである。

問 次の言葉から連想される言葉をできるだけたくさん書いてください。

キーワード例 ディナー

解答例 夕食 レストラン ステーキ

4.2.2 連想概念の記述タスク:実験結果

タスク 2 に参加したワーカーはのべ 235 人で、平均実行タスク数は 4.3 タスクであった。獲得できた連想語の数は、1 キーワードあたり約 30 語であり、全体では約 3000 語が獲得できた。チェック問題に関しては 89.3% のワーカーが正しく回答しており、ほとんどのワーカーが真剣に作業に取り組んだと考えられる。獲得された連想概念については図 3 に例を示す。クラウドソーシングで得られた連想概念の右の数値は、同じ回答をしたワーカー数であ

る。右のカラムには参考として自動獲得された連想概念の例をあげた。自動獲得された連想概念の右の数値は関連度である。例えば「蜂蜜」などは自動獲得の結果はさほど悪くはなく、クラウドソーシングで獲得された連想概念と非常に近いものが獲得されている。一方で「直線」に関しては、自動獲得の場合は「馬」や「追走」などおそらく「競馬」に関連すると思われる単語が獲得されている。クラウドソーシングを利用した場合は「定規」や「1 次関数」などが連想概念として獲得されており、自動獲得の結果よりよいものが獲得されていることがわかる。

4.3 予備実験のまとめ

実験結果から、連想概念のチェック、連想概念の記述に関しては特に問題なく行うことができることがわかった。この結果から、自動獲得された連想概念の質を、ゲーミフィケーションにより正しくチェックできるのであれば、その語の修正は高い質で行うことができると考えられる。また、自動獲得結果の中にはクラウドソーシングで獲得された連想概念と同程度の内容のものが含まれており、連想ゲームとして十分に機能することが期待される。

5 まとめと今後の展望

本稿では、自動的な手法、クラウドソーシング、ゲーミフィケーションを効果的に組み合わせることで、大規模で質の良いリソースを安価に構築する枠組みについて論じた。また、クラウドソーシングを利用して、関連語の部分的な修正を高い精度で行うことが可能であることを示した。現状ではまだ、実際のゲームログを得るには至っていないが、今後、実際にゲームログが獲得できれば、さらなる分析を進めたい。本発表の場では、実際のログに基づいた考察を行いたいと思っている。

参考文献

- [1] Collin F. Baker, Charles J. Fillmore, and John B. Lowe. The berkeley framenet project. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 1*, ACL '98, pp. 86–90, Stroudsburg, PA, USA, 1998. Association for Computational Linguistics.
- [2] Chris Callison-Burch. Fast, cheap, and creative: Evaluating translation quality using Amazon's Mechanical Turk. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language*

表 2 自動獲得連想概念のチェック例

キーワード	自動獲得された連想概念 (下線はワーカーによる選別で連想概念とされたもの)
贈り物	引き出物, 一杯だ, 御礼, 誕生日, クリスマス, ラッピング, 沢山だ, 御返し, 最適だ, 還暦, 誕生日プレゼント, 結婚, 記念日, ホワイトデー, 感謝, 盛り沢山だ, ムトウ, 母の日, 出産, ギフト, 記念品, 記念, 心, <u>プレゼント</u> , 理想的だ, 敬老の日, 歳暮, 中元, 相応しい, 一般的だ, 内祝い, 神様, 父の日
王国	中古バイク, 建国, 国, 王子, フルーツ, 勢力, 楽天市場店, 国王, 雑貨, 情報検索サイト, 王, 帝国, 支配, 買取, 価格情報, 好きだ, ヤフー店, 嬉しい, 首都, 軍, 王女, ホンダ

表 3 自由記述によって得られた連想概念の例

キーワード	クラウドソーシングで得られた連想概念	自動獲得された連想概念
蜂蜜	ミツバチ:5, ローヤルゼリー:3, ハニー:2, 甘い:2, 蜂:2, ...	クローバー:0.9, 砂糖:0.81, 蜜蜂:0.58, 養蜂:0.58, 蜜:0.45
直線	定規:6, 道路:2, 曲線:2, 直線道路:1, 1次関数:1, ...	馬:0.9, 曲線:0.73, 番手:0.45, コーナー:0.4, 追走:0.36

Processing, pp. 286–295, Singapore, August 2009. Association for Computational Linguistics.

- [3] Seth Cooper, Adrien Treuille, Janos Barbero, Andrew Leaver-Fay, Kathleen Tuite, Firas Khatib, Alex Cho Snyder, Michael Beenen, David Salesin, David Baker, and Zoran Popovic. The challenge of designing scientific discovery games. In *FDG*, pp. 40–47, 2010.
- [4] Christiane Fellbaum. *WordNet: An Electronic Lexical Database*. Bradford Books, 1998.
- [5] Marco Fossati, Claudio Giuliano, and Sara Tonelli. Outsourcing framenet to the crowd. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 742–747, Sofia, Bulgaria, August 2013. Association for Computational Linguistics.
- [6] Daisuke Kawahara, Yuichiro Machida, Tomohide Shibata, Sadao Kurohashi, Hayato Kobayashi, and Manabu Sassano. Rapid development of a corpus with discourse annotations using two-stage crowdsourcing. In *Proceedings of the 25th International Conference on Computational Linguistics (COLING2014)*, Dublin, Ireland, 2014.8.
- [7] Mathieu Lafourcade. Making people play for lexical acquisition. In *Proceedings of the 7th Symposium on Natural Language Processing (ENLP07)*, pp. 13–15, Pattaya, Thailande, December 2007.
- [8] Rion Snow, Brendan O’Connor, Daniel Jurafsky, and Andrew Y. Ng. Cheap and fast—but is it good?: Evaluating non-expert annotations for natural language tasks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP ’08, pp. 254–263, Stroudsburg, PA, USA, 2008. Association for Computational Linguistics.
- [9] Daniele Vannella, David Jurgens, Daniele Scarfini, Domenico Toscani, and Roberto Navigli. Validating and extending semantic knowledge bases using video games with a purpose. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1294–1304, Baltimore, Maryland, June 2014. Association for Computational Linguistics.
- [10] Noortje J. Venhuizen, Valerio Basile, Kilian Evang, and Johan Bos. Gamification for word sense labeling. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Short Papers*, pp. 397–403, Potsdam, Germany, March 2013. Association for Computational Linguistics.
- [11] Luis von Ahn. Games with a purpose. *IEEE Computer*, Vol. 39, No. 6, pp. 92–94, 2006.
- [12] Omar F. Zaidan and Chris Callison-Burch. Crowdsourcing translation: Professional quality from non-professionals. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*, HLT ’11, pp. 1220–1229, Stroudsburg, PA, USA, 2011. Association for Computational Linguistics.
- [13] Naomi Zeichner, Jonathan Berant, and Ido Dagan. Crowdsourcing inference-rule evaluation. In *Pro-*

ceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 156–160, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

- [14] 河原大輔, 黒橋禎夫. 格フレーム辞書の漸次的自動構築. *自然言語処理*, Vol. 12, No. 2, pp. 109–131, mar 2005.
- [15] 進義治, 黒橋禎夫. 名詞関連語知識に基づく文章のグラフ表現とその応用. *言語処理学会 第 20 回年次大会*, pp. 1007–1010, 札幌, 2014.3.20.