

# 英文閲覧時における視線と音声情報を組み合わせた対訳提示手法

岡田友哉<sup>†1</sup> 坂本大介<sup>†1</sup> 小野哲雄<sup>†1</sup>

**概要:** 英語文書閲覧時にユーザへの負担が大きくなるような対訳提示手法として視線入力と音声情報を組み合わせた対訳提示手法を検討する。具体的には、視線をポインティングに利用し、対訳提示のトリガーとして音声入力を用いることで、スムーズな対訳提示を目指す。対訳提示の方法として視覚情報による提示と音声情報による提示を実装し、既存のアイデアである視線のみを用いた対訳提示を含めた3手法の比較評価を行なった。評価実験の結果、操作時間に関しては視線のみを用いた対訳提示に劣るが、ユーザビリティ評価では大きな差は確認されなかった。また、音声を用いることの利点として、音声入力の利用による *Midas touch problem* の解消や、対訳を読み上げることで文章上に対訳を表示する必要がなく、閲覧の妨げになりにくいといった利点が確認された。

## 1. はじめに

近年の国際化に伴い、英語を母語としない日本人が日常的に英語を目にする機会が増えている。Web サイト上で使用されている言語の割合は、英語が 50%を超える一方で日本語は 3%程度に留まり[1]、英語の使用率が非常に高い。しかし、私たち日本人にとって非母語である英語を完全に理解するのは容易ではなく、翻訳ツールなどを用いて英語の読解を進める人が多いのが現状である。英文中の単語の日本語訳を得るための方法として、しばしば翻訳サイトや、マウス操作で選択することでポップアップとして日本語訳を表示するツールが用いられるが、学术论文のような長文を読む際には操作を繰り返し行うことでユーザへの負担が増大することが予想される。また、近年の機械翻訳技術の発展により、機械翻訳の精度が向上していることは確かであるが、長文や複雑な構造の文章に対する精度は満足できるものとは言えず、未だ人間による解釈が必要だと言える。

本研究では、ユーザが英文中の単語訳を得る際の負担が小さくなるような対訳提示手法を検討する。ユーザの操作を可能な限り少なくするために、今回は視線を用いた操作に注目する。視線を用いたユーザインタフェースは多く研究されており、手を自由に使うことができない状況に向けて提案されることが多い。そのため、対訳を提示するための操作として利用することで、ハンズフリーな操作が可能となり、ユーザが対訳を得る際の操作やその負担の減少が期待される。

視線情報を用いた辞書引きシステムを提案した先行研究[2]は存在するが、視線情報のみを用いていることから *Midas touch problem*[3]の影響を受けやすいという問題点があった。*Midas touch problem* とは、操作としての視線の動きと、それ以外の視線の動きが区別しにくいという問題である。これを解消するために様々な手法が提案されており、手法の1つとして音声など他の入力モダリティと組み合わせる手法が提案されている[4][5][6]。本研究では、視線情報と音声入力を組み合わせることで *Midas touch problem* を解

消し、ユーザへの負担が小さくなるような対訳提示手法を検討する。視線と音声の双方を利用した手法として2つの手法を実装した。これらは共通して視線情報をポインティングに用い、音声入力を対訳提示のトリガーに用いており、2つの手法は対訳の提示方法が異なる。1つ目は視覚情報としてポップアップで対訳を提示し、2つ目は音声情報として対訳を読み上げて提示する。これらの手法と視線情報のみを用いた手法の全3手法の比較評価を行い、視線を用いた対訳提示における音声の有用性について検証した。

## 2. 関連研究

視線を用いてコンピュータの操作を行う時、大きな問題となるのが *Midas touch problem*[3]である。*Midas touch problem* による意図しない動作の発生を防ぐため、多くの手法が提案されている。問題の解決のために一定時間の注視がしばしば用いられる。Hansen らは注視により操作の決定を行う *Eyecon* を開発し、その注視時間は 0.5 秒が最適だと述べている[7]。また、他の入力モダリティと組み合わせることで *Midas touch problem* を解消する手法は多く提案されている。Pfeuffer らはタブレット端末を保持している時の親指でのタッチ操作を、視線によるポインティングの補助や操作の決定に用いた[4]。Vieira らは視線と音声入力を組み合わせることで、ユーザの視線が示す対象を、音声入力中の指示語の対象に対応付けるブラウザを実装した[6]。

文書閲覧という状況に注目した研究も存在する。吉村らは装着型のアイトラッカーを用いて、fixation の持続時間や saccade 長などを特徴量とし、SVM (Support Vector Machine) を用いた文章の理解度推定を行なった[8]。大社らは読みの速さや読み返しの回数、fixation の持続時間を用いて、ユーザが難しいと感じた部分を推定し、その情報の可視化を行なった[9]。また、非母語の意味理解に関する研究もされており、平岡らは視線情報に関する特徴量と英単語の語彙頻度に関する特徴量から、SVM を用いて未知語の検出を行なった[10]。

<sup>†1</sup> 北海道大学

本研究と関連が強い、視線を用いた非母語の対訳の取得に関する研究も存在する。東中らは視線情報や文書中の英単語に関する情報などを特徴量とし、決定木学習を用いてユーザが英単語訳を必要としているかどうかの推定を行ない、自動辞書引きシステムを実装した[2]。他言語では、Hyrskykari らはユーザの読書中の問題を検出し、辞書引きを行う iDict を開発した[11]。しかし、東中らのシステムも含め、多くの先行研究では視線からユーザ状態を推定することに重きが置かれており、実際の使用を考えた時に重要となるユーザビリティに関する議論は十分にされていない。近年は視線検出が手軽にできるようになってきているため、視線を用いたインターフェースの実用化が現実的になっており、ユーザビリティ評価を行う必要性が高まっている。また、東中らのシステムではシステムが自動で辞書引きを行うため、ユーザが意図して辞書引きを行えない点が問題点として挙げられており、ユーザが主体的に辞書引きを行えるような手法を考える必要がある。Ho らは視線を利用して英文の訳を表示するインターフェースを実装し、訳の表示方法について調査を行っており、ユーザが主体的に操作を行うことの利点を述べている[12]。そこで本研究では、ユーザが主体的に対訳を得ることができ、さらに Midas touch problem を解決することができる手法として、視線によるポインティングと音声入力による対訳提示のトリガーを用いた手法を検討する。また、ユーザビリティ評価を行い、視線と音声情報を入力に用いた 2 手法と視線のみを入力に用いた手法の全 3 手法について比較を行なった。

### 3. 提案手法

提案手法では、まず、ユーザは視線を用いて対訳を得たい単語の選択を行う。対訳を得たい英単語を見たまま、音声入力を行うことで選択された単語の対訳が提示される。従来の視線を用いた対訳提示を行う手法では、視線の動きからユーザが対訳を必要としているかどうか推定していたが、本手法では音声入力をトリガーとして用いることで、ユーザが主体的に、かつ簡単に対訳を得られることを目標としている。対訳の提示方法としてポップアップを用いた視覚情報による提示と、音声で読み上げる音声情報による提示の 2 種類の方法を実装した。

#### 3.1 インタラクションデザイン

提案手法について、具体的な操作方法を説明する。以降、視線と音声情報を入力とし、対訳をポップアップにより提示するものを手法 VP (Voice-Popup)、視線と音声情報を入力とし、対訳を音声で読み上げることで提示するものを手法 VV (Voice-Voice) とする。

##### ・手法 VP (Voice-Popup)

本手法では、ユーザは対訳を知りたい英単語を見ながら

「辞書」、「翻訳」、「意味」、「訳」のいずれかを音声で入力することで、対象の英単語の対訳が図 1 のようなポップアップで表示される。対訳の確認が終わった後は 0.5 秒以上ポップアップの範囲外を見るかポップアップを見続けることでポップアップが消去され、次の入力が可能になる。



図 1 ポップアップによる対訳の表示例

##### ・手法 VV (Voice-Voice)

本手法では、対訳を知りたい英単語を見ながら音声入力を行うと、対象の英単語の対訳が英語、日本語訳の順で読み上げられる。入力方法に関しては手法 VP と同様であるが、対訳の提示方法が音声情報を用いたものとなっている。対訳の読み上げが終了し次第、次の入力が可能となる。

#### 3.2 使用機器

今回、ユーザの視線情報を取得するアイトラッカーには Tobii EyeX を用いた。Tobii EyeX は角膜反射法により視線検出を行うアイトラッカーであり、サンプリングレートは 60Hz である。操作に用いる PC には Intel® Core™ i5-4200U を搭載した Fujitsu Lifebook SH90/M (ディスプレイサイズ: 13.3 型ワイド, 解像度:2560×1440) を用いた。音声入出力には使用した PC に搭載されているマイク、スピーカーを用いた。

#### 3.3 実装

提案手法の実装について説明する。ユーザが実際に見ている位置と検出された視線の画面上の座標が一致しているかを判断するため、SDK で標準搭載されている視覚的なフィードバックを使用した。英単語の選択には視線によるポインティングを用い、検出される視線の画面上の座標と表示されている英単語の一致判定を行うことで、選択対象の英単語の決定を行う。一致判定は、単語の幅と、単語に含まれるアルファベットの最大の高さからなる矩形領域を単語領域とし、その中に視線座標が含まれているかどうかで判定を行う。各英単語の対訳の取得は翻訳 API などを用いることでリアルタイムな取得が可能であるが、今回は簡単な実装とし、予め対訳を保持しているものとした。音声入力部分では、Windows10 標準搭載の音声認識エンジンを用いて、登録された言葉が認識された時に入力を受け付ける。今回は、音声認識の精度上、ユーザによって言葉が認識されにくい場合があるため、前述した 4 つの言葉を用意し、その中から任意で使用させた。手法 VV における音声出力には、Windows10 標準搭載の音声合成エンジンを用いた。

## 4. 評価実験

提案手法の評価のため、2つのタスクを用いた参加者実験を行った。また、アンケート調査やインタビューを行い、ユーザビリティに注目した評価を行なった。実験を実施する際には、参加者の承諾を得た上で実験風景を映像として記録した。目を意識的に動かす内容の実験であるため、参加者の疲労を考慮し休憩を促しながら実験を行なった。

### 4.1 実験手順

本実験には、大学生および大学院生7名（男性6名、女性1名、平均23.6歳、標準偏差2.2）が参加した。そのうち、2名が実験時に眼鏡をかけていた。7名のうち、1名が視線を入力として使用した経験があった。参加者毎にタスクを行う前にアイトラッカーのキャリブレーションを行い、正確に視線の検出ができていないか確認しながらタスクを実施した。参加者には視線が正確に検出できていない場合は申告するように伝え、必要に応じて再度キャリブレーションを行なった。

次に実験の手順を説明する。初めに参加者に簡単なアンケートに答えてもらい、続いて実験の概要説明、アイトラッカーのキャリブレーション、タスクの説明を行った。説明の終了後、3章で述べた2つの手法と、比較のための視線のみを用いた操作手法について、それぞれ2つのタスクを実施した。タスクは手法毎に、操作方法の説明、練習、タスク1、タスク2、アンケートという流れで実施した。各タスクの詳細については次節に記述する。なお、タスクを行う手法の順番は参加者毎にランダムである。全ての手法についてタスクを終えた後、実験全体に関するアンケートに回答してもらった。

### 4.2 実験タスク

本実験では、3章で述べた2つの手法と、比較のための視線のみを用いた操作手法について、それぞれ2つのタスクを実施した。比較のため実装した、視線のみを用いた手法では、ユーザが対訳を得たい単語を0.5秒間見続けると、ポップアップにより対訳が提示される。以降これを手法GP（Gaze-Popup）とする。各手法について、タスクを行う前にタスク1を模した練習を行い、操作に慣れさせてからタスクを行なった。実験の様子を図2に示す。

#### ・タスク1

1つ目のタスクでは、画面上に表示されている英文から、システム側で指定した英単語の対訳を表示させるタスクを行なった。ここで使用する英文は計70語程度の英文であり、Wikipediaよりランダムで記事を選び、その記事中から選択した。各手法で使用される英文は異なるものを用いており、手法と英文の組み合わせは参加者毎にランダムで決定



図2 実験の様子

Sofia Gruskin is a scholar and advocate in the field of health and human rights whose contributions range from global policy to the grassroots level. For more than 25 years her work has been instrumental in developing the conceptual, methodological, and empirical links between health and human rights, with a focus on sexual and reproductive health, HIV and AIDS, child and adolescent health, gender-based violence, non-communicable disease, and health systems.

図3 タスク1における英文の表示例

している。対訳を表示する英単語は文中からランダムで一つ選択し、単語の背景色を黄色に変更することで指定した。指定した英単語の対訳の表示は15回繰り返し行われた。タスクに制限時間は設けておらず、参加者が誤った単語を選択してもタスクは継続し、15回全て成功するまで終了しない。このタスクでは各参加者について、各対訳表示の完了時間とタスク中の失敗回数を記録した。本タスクにおける英文の表示例は図3に示す。

#### ・タスク2

2つ目のタスクでは、実際の利用シーンを想定してユーザビリティ評価を行うため、画面上に表示した英文の閲読を行ってもらった。この際システムの使用は自由とし、出来るだけ英文の意味を理解するように指示した。閲読の終了タイミングは参加者の任意とした。ここで使用した英文は、タスク1とは異なる英文であり、システムの使用を促すため、読解の難易度が高くなるように心理学系の論文[13][14][15]から80語程度の英文を選び使用した。手法と英文の組み合わせはタスク1と同様にランダムで決定している。

### 4.3 計測方法

各手法について、タスク1における対訳表示完了時間を計測した。これはタスク1において、システムが対訳を表示する英単語を指定してから、その英単語の対訳の表示が終了するまでの時間である。具体的には、手法VPおよびGPでは、対象の英単語の背景色が黄色に変化してから、そ

の英単語の対訳が示されたポップアップを表示後、消去されるまでの時間を指し、手法 VV では、対象の英単語の背景が黄色に変化してから、その英単語の対訳の読み上げが終了するまでの時間を指す。

また、各手法についてタスク終了後に System Usability Scale (SUS)[16]と自由記述からなるアンケートに回答してもらい、口頭で気付いた点などを尋ねた。全ての手法のタスク終了後、実験全体について感想などを記述してもらった。

## 5. 実験結果

### 5.1 対訳表示完了時間

各手法について、タスク 1 におけるそれぞれの対訳表示完了時間の平均と標準誤差をまとめたものを図 4 に示す。手法 VP は  $5.48 \pm 0.46$  秒、手法 VV は  $9.35 \pm 0.83$  秒、手法 GP は  $3.69 \pm 0.38$  秒という結果になった。この結果に対して 1 要因分散分析（参加者内計画）を実施したところ、有意な差が確認された [ $F(2,12) = 32.94, p < .01$ ]。ここで、Holm 法を用いた多重比較を実施すると、全ての手法の間で有意な差が確認された（全て  $p < .01$ ）。

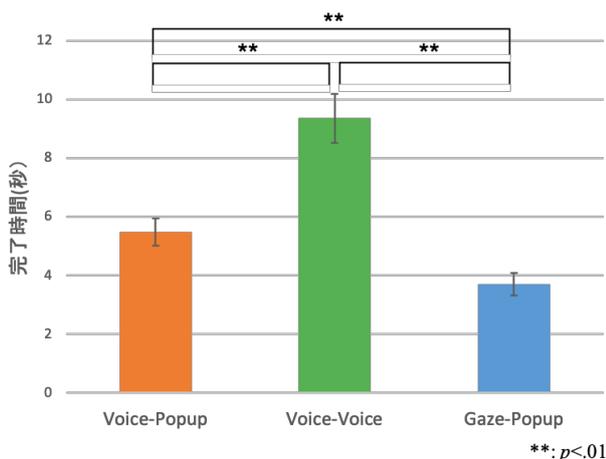


図 4 タスク 1 における対訳表示完了時間

### 5.2 ユーザビリティ

各手法についての SUS の結果を図 5 に示す。SUS は 5 段階のリッカート尺度の質問 10 項目からなるアンケートであり、各質問項目の評価値を基に 100 点満点のスコアを算出することができる。今回の参加者の平均値および標準誤差は図 5 のような結果となっており。手法 VP は  $64.29 \pm 5.45$  点、手法 VV は  $65.36 \pm 6.23$  点、手法 GP は  $69.29 \pm 5.79$  点であった。若干ではあるが、手法 GP が他の 2 手法に比べてユーザビリティの評価が高い傾向が見られる。

また、各質問項目に対する回答の平均値は図 6 のようになった。SUS は奇数番号の質問項目はポジティブな内容の質問であり、偶数番号はネガティブな内容の質問である。

質問項目全体を通して、手法 GP, VV, VP の順で高い評価となる傾向が見られる。Q7「ほとんどの人がこのシステムの使い方を簡単に学べると思う」については他の質問項目に比べて評価が高くなる傾向があり、3 手法共通して学習コストがあまり高くないという結果となっている。

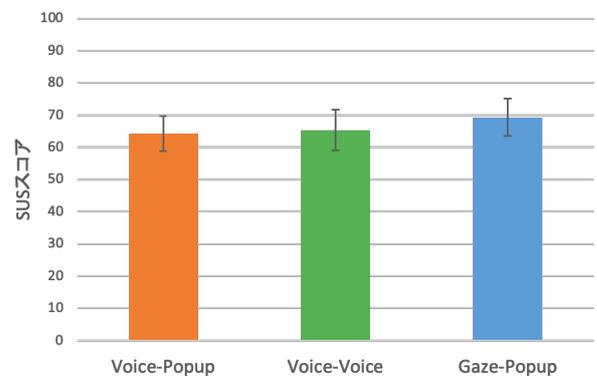


図 5 SUS スコア

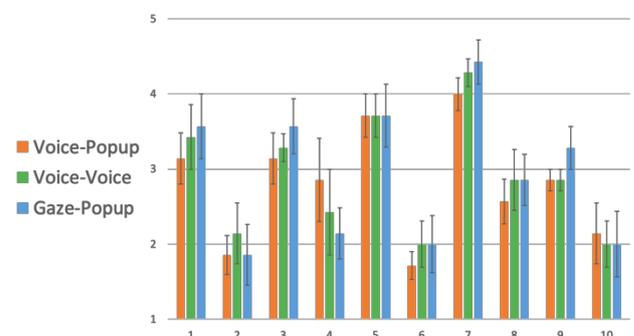


図 6 SUS 各質問項目の結果

## 6. 議論

タスク 1 で得られた対訳表示完了時間から、手法 VV は他の 2 手法に比べて操作を行う時間が長くなることがわかった。また、手法 VP に関しても手法 GP に比べて操作時間が長い。このような結果になった原因として、音声入力による入力時間の増加が考えられる。今回の設定では、手法 GP は 0.5 秒間の注視により対訳の提示を行うが、音声入力を加えることで 0.5 秒以上の入力時間がかかってしまった可能性がある。参考までに手法 VP のタスク 1 における音声認識が行われた際の注視時間の平均値を算出したところ、およそ 0.98 秒という結果が得られた。また、手法 GP に関するアンケートの自由記述欄への回答に「音声の誤認識によるロスがないのが良い」という記述があり、音声認識の精度も対訳表示完了時間へ影響を与えていることが示唆される。また、手法 VP と手法 VV の間の完了時間の差は音声による読み上げが主な原因となっていることが考えられる。

タスク 1 における完了時間に大きな差が確認された一方で、ユーザビリティ評価に関しては評価値の差があまり大

きくない傾向が見られた。特に手法 VP, 手法 VV 間の SUS スコアの差は小さく, 完了時間以外の観点で評価された可能性がある。手法 GP に関する意見として「ポップアップが文と重なり邪魔になる」というものがあり, これは手法 VP に関しても同様のことが言える。今回は英文を読むことに注目しているため, それを妨げるような動作は大きな問題となり得る。音声による提示の場合はこのような問題が起こることがないため, その点で評価された可能性がある。また, 手法 GP に対して「意図しないところでポップアップが出ることもある」という意見があったが, 手法 VP では「意図しない注釈があまり出なくて良かった」という意見があり, Midas touch problem が解消できていることも確認できる。

一方で, 音声インタフェース特有の問題点も確認された。音声を入力とすることに対し, 「人前では使いたくない」という意見や, 音声を出力に用いる場合に, 「音声聞き取りにくいことがある」, 「同音異義語があると正しい意味が分からない」という意見があった。音声インタフェースの問題点として, ユーザの周辺環境に影響を受けることが挙げられる。入力に関しては, 単に他人がいる前での使用に抵抗が生まれるだけでなく, 周囲の騒音などに影響を受けて正しい入力がされない場合もある。また, 出力についても同様に環境音により阻害される可能性は高い。以上から, 音声情報を対訳提示に用いる際, 使用する状況がある程度限られることが分かった。一方でポジティブな意見として, 「英語の学習用に使えたら楽しそう」, 「障がいを持った人への補助に使えそう」という意見が得られ, 使用目的をある程度絞ることで, 提案手法がより有用になる可能性が示唆された。最後に, 提案手法の使いやすさには視線検出, 音声認識両方の精度が大きく影響を与える。共に十分な精度が保証された上での評価, 実用が必要になるだろう。

## 7. まとめ

本稿では, 英語文書を閲覧する際にユーザへの負担が大きくなりたくないような意味理解支援を行うことを目的として, 視線と音声を組み合わせた対訳提示手法を提案した。視線をポインティング, 音声入力を対訳提示のトリガーとして用い, 提示方法として視覚情報を用いた手法と音声情報を用いた手法の2種類を実装し, 視線のみを用いた手法を含めた3手法について比較実験を行なった。その結果, 操作時間としては視線のみの手法の方が短かったが, ユーザビリティ評価は若干視線のみの手法が高い傾向はあるものの, 大きな差は確認されなかった。音声を入出力に用いた手法は Midas touch problem の解消や文書閲覧の妨げになりにくいことなど操作時間以外の観点で有用である可能性があり, 学習支援や障がいを持った人への支援といった目的での利用が期待できる。

## 参考文献

- [1] “Usage Statistics and Market Share of Content Languages for Websites, December 2018”  
[https://w3techs.com/technologies/overview/content\\_language/all](https://w3techs.com/technologies/overview/content_language/all), (参照 2018-12-4).
- [2] 東中竜一郎, 大野健彦. 視線を用いた自動辞書引きシステム. 言語処理学会第9回年次大会併設ワークショップメディア/モダリティ統合における言語処理, Vol. 13, 2003.
- [3] Robert J. K. Jacob. The use of eye movements in human-computer interaction techniques: What you look at is what you get. ACM Trans. Inf. Syst., Vol. 9, No. 2, pp.152-169, 1991.
- [4] Ken Pfeuffer and Hans Gellersen. Gaze and touch interaction on tablets. In Proc. UIST '16, ACM, pp. 301-311, 2016.
- [5] T. R. Beelders and P. J. Blijnaut. Using eye gaze and speech to simulate a pointing device. In Proc. ETRA '12, ACM, pp.349-352, 2012.
- [6] Diogo Vieira, João Dinis Freitas, Cengiz Acartürk, António Teixeira, Luís Sousa, Samuel Silva, Sara Candeias, and Miguel Sales Dias. “read that article”: Exploring synergies between gaze and speech interaction. In Proc. ASSETS '15, pp. 341-342, 2015.
- [7] John Paulin Hansen, Allan W Andersen, and Peter Roed. Eye-gaze control of multimedia systems. Advances in Human Factors/Ergonomics, Vol. 20, pp. 37-42, 1995.
- [8] 吉村和代, 川市仁史, 黄瀬浩一. アイトラッカで取得した視点情報と文書理解度の関係 (テーマセッション, 文字・文書の認識と理解及びアルゴリズム・システム評価). 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 112, No. 495, pp. 261-266, 2013.
- [9] 大社綾乃, Kai Kunze, Olivier Augereau, 黄瀬浩一. 学習補助のための視点情報に基づく文書アノテーション (信号処理). 電子情報通信学会技術研究報告= IEICE technical report: 信学技報, Vol. 115, No. 22, pp. 161-166, 2015.
- [10] 平岡類, 田中宏季, サクティサクリアニ, 吉野幸一郎, ニュービッググラム, 中村哲. 注視特徴を用いた svm による非母語読解時の未知語検出 (ヒューマンコミュニケーション基礎). 電子情報通信学会技術研究報告= IEICE technical report: 信学技報, Vol. 116, No. 31, pp. 213-218, 2016.
- [11] Aulikki Hyrskykari, Päivi Majaranta, Antti Aaltonen, and Kari-Jouko Räihä. Design Issues of iDict: A Gaze-Assisted Translation Aid. In Proc. ETRA '00, pp. 9-14, 2000.
- [12] Tien-Yu Ho, Hao-Chuan Wang, and Shong-Hong Lai. Non-native Language Reading Support with Display of Machine Translation Based on Eye-Tracking and Sentence-Level Mapping. In Proc. ChineseCHI '18. pp. 57-63, 2018.
- [13] Ze Zhu, Jian Li, Bo Zhang, Ye Li, and Houcan Zhang. The effect of motivation and positive affect on ego depletion: Replenishment versus release mechanism. International Journal of Psychology, Vol. 52, No. 6, pp. 445-452, 2017.
- [14] Magdalena A. Żemojtel-Piotrowska, Jarosław P. Piotrowski, and John Maltby. Agentic and communal narcissism and satisfaction with life: The mediating role of psychological entitlement and self-esteem. International Journal of Psychology, Vol. 52, No. 5, pp. 420-424, 2017.
- [15] Philippe Bou Malham and Gerard Saucier. The conceptual link between social desirability and cultural normativity. International Journal of Psychology, Vol. 51, No. 6, pp. 474-480, 2016.
- [16] John Brooke. SUS- A quick and dirty usability scale. Usability evaluation in industry, pp.189-194, 1996.