

英語映画の非発話情報に付与された字幕に関する分析

福 島 孝 博

1 はじめに

映画に付与される字幕には、大別して2種類ある。1つは通常の子幕であり、健聴者が視聴するものである。もう1つは、難聴者用字幕で、映画の登場人物の発話だけでなく、その他の音声情報も字幕にしているものであり、北米においては、SDH (Subtitles for Deaf and Hard-of-hearing) と呼ばれており、多くの映画が媒体 (DVD、Blu-ray など) として販売される際に付与されている。

映画には、登場の人物が話す発話の音声があるが、それに加えて、映像にはそれ以外の音声も多く含まれている。例えば、明確には聞き取れない人のひそひそ話、笑い声、動物の鳴き声、機械の音、効果音などである。これらを発話以外の情報として「非発話情報」(Non-Speech Information, NSIと略す) とし、それに対する字幕を非発話情報の字幕とする。非発話情報の字幕は、通常の子幕にはあまりないが、難聴者用字幕、SDHにおいては数多く出現する。これらは当然ながら、映像にある発話以外の音声情報を字幕化して耳の不自由な方の理解を助けるものである。

映画やテレビ番組での字幕、特に、非発話情報の字幕についての研究は、まだ、数少ない。Zdenekは、その著書で、サイエンスフィクション分野の4本の映画について調査をし、それらに出現する非発話情報字幕を定量的、定性的に分析をしている ([1])。また、福島は、同一映画に付与された日本語と英語の非発話情報字幕を比較、分析しているが、調査の対象が1本の映画 (ハリウッド版Godzilla) の字幕だけであり、小規模なものとなっている ([2] [3])。

本稿では、北米制作の5本の英語の映画において付与されている非発話情報に対する英語での字幕を調査対象として、その頻度や、内容による分類、映画間の比較などを通じて分析をしている。また、人工知能の一分野である自然言語処理で注目されているword2vecを使い、字幕を単語分散表現での類似度の点から分析をしている。

以下に調査の対象、手法、結果を示し、その後に結果についての分析およびword2vecでの分析を述べる。

2 調査対象

調査の対象とした映画は以下の5本である。これらは、比較的最近の映画であり、そのジャンルを幅広くし、ミステリー、アクション、サイエンスフィクション、子ども向け映画となっている。

発話を誰が行ったかを示す発話者字幕は、本来非発話情報の字幕に含まれるが、今回の調査の対

象から外し、それ以外の非発話情報に対する字幕を対象としている。

表 1 調査対象の映画

タイトル	制作年	時間 (分)	媒体
Dark Shadow	2012年	113	Blu-ray
Oblivion	2013年	125	Blu-ray
The Tourist	2010年	103	Blu-ray
Boss Baby	2017年	97	Blu-ray
Godzilla (ハリウッド版)	2014年	123	Blu-ray

3 調査手法

各映画を再生し、付与された字幕から非発話情報に対して付けられた字幕を人手で書き起こした。また、それらの字幕を内容によって分類した。表2に分類を示している。分類は、Zdenekの分類に基づいて行った ([1])。表中の字幕例は、調査対象の1つであるDark Shadowに出現したものである。

表 2 非発話情報の字幕の分類

区分 (省略形)	内 容	字 幕 例
Language identifiers (LI)	言語を示す (外国語の場合)	(SPEAKERS IN FRENCH)
Sound effects (SE)	効果音	(THUNDER CRACKING) (MACHINE WHIRRING)
Paralanguage (Para)	パラ言語 人の音声で明確な発話とされないもの	(GRUNTING) (INAUDIBLE DIALOGUE) (ALL LAUGHING)
Manner of speaking (MS)	発話の様子を示す (その直後に発話がある)	(SPEAKING INDISTINCTLY)
Music (Music)	音楽	♪～ (SYNTHESIZER PLAYS MELLOW MUSIC)
Channel identifiers (CI)	音源情報	(OVER RADIO)

尚、1つの字幕が複数の区分に該当する場合があった。例えば、Touristに出現する (MAN SPEAKING ITALIAN ON PA) という字幕の場合、イタリア語であるのでLI (Language Identifiers) の字幕であり、且つ、ON PA、PAとはPublic Addressの略、「拡声装置で」という音源情報を含んでいるので、CI (音源情報) とも分類される。

4 調査結果

まず、各作品に現れた非発話情報の字幕数、その平均値と標準偏差値は以下のとおりとなる。

表3 非発話情報の字幕数、平均値と標準偏差値

映画タイトル	NSI字幕数
Dark Shadow	104
Oblivion	86
Tourist	131
Boss Baby	362
Godzilla	310
平均	198.6
標準偏差	127.8

NSIの字幕数は、映画の作品によってばらつきがあり、最低は86から最高は362である。その為、標準偏差が大きな値となっている。

次に、非発話情報字幕を表2に示された6区分に分類すると以下のようになる。

表4 区分別の非発話情報字幕数

Movie title	LI	SE	Para	MS	Music	CI
Dark Shadow	1	25	62	2	12	10
Oblivion	0	17	60	3	5	1
Tourist	75	29	19	0	3	8
Boss Baby	3	45	298	1	8	8
Godzilla	28	142	115	3	0	39
Total	107	258	554	9	28	66

全体的にパラ言語の字幕が多くあり、次に効果音の字幕が多くなっている。また、映画によっては、その内容から外国語であるという字幕、LIが多くなっているものがある。

5 分 析

5本の映画の非発話情報字幕について、まず、全体的な傾向を述べ、その後に、多くの字幕が分類されたパラ言語、効果音の字幕について分析する。その後に、自然言語処理の分野で使われているword2vcを使い、パラ言語の字幕の分析をしている。

5.1 NSI字幕の傾向

全体としては、パラ言語や効果音に区分される字幕が多くなっている。図1では、映画ごとに100%になるように区分された字幕の割合を示している。但し、映画ごとに違いもみられる。

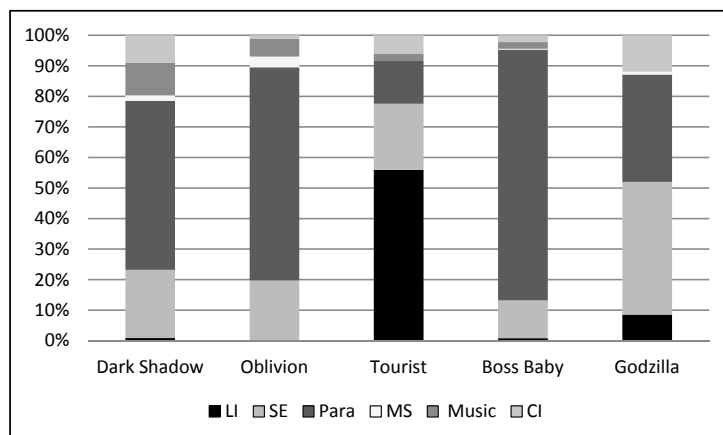


図1 非発話情報字幕の区分別の割合（100%表示）

Dark ShadowやGodzillaではCI（音源情報）の字幕が比較的多くなり、TouristではLI、言語を示す字幕が相当に多くなっている。また、パラ言語については、3本においてその割合が一番多くなっているが、TouristとGodzillaではそうではない。Boss Babyでは、圧倒的にパラ言語の字幕が多くなっており、その数も他の映画と比べて多くなっている。Godzillaにおいては、SE、効果音が多くなっている。

これらは、映画の内容を反映したものであり、Touristでは、映画の設定場面がヨーロッパであり、外国語で話される発話については、何語であるのかを字幕で示している。また、Boss Babyでは、多くの登場人物がいろいろな動作をするため、パラ言語と区分される字幕が多くなっていると思われる。Godzillaにおいては、ゴジラの咆哮が何度も繰り返されるが、これがSEと分類されるため、SEの字幕が増えている。

5. 2 パラ言語字幕

パラ言語に対する字幕については、英語に多く出現し、日本語では少ないとの指摘が既にあるが（[2]）、ここでは、英語でのパラ言語に対する字幕の特徴を示す。

まず、パラ言語字幕の総数（出現延べ数）は、554であり、種類数（異なり数）109である。

次に、調査対象全体において出現するパラ言語字幕について、高頻度（10回以上）のものをまとめると以下ようになる。

表5 高頻度のパラ言語字幕

字 幕	頻度	字 幕	頻度	字 幕	頻度
GASPS	37	SCREAMING	25	PANTING	17
GRUNTING	32	SIGHS	25	EXCLAIMING	14
GRUNTS	31	CRYING	20	CHATTER	10
LAUGHING	28	CHUCKLES	19		
YELLING	28	LAUGHS	18		

また、その他の10回未満の語をみると、GASPS, GASPINGのように派生語関係となっている語がいくつか含まれている。これらの派生語、具体的には3種類で、原形、原形 + s、原形 + ingとなっている。それらをまとめると以下ようになる。

表6 派生形を含めた高頻度のバラ言語字幕

Expressions	Frequency	Total Frequency	Expressions	Frequency	Total Frequency
GRUNT	1	64	SIGH	1	26
GRUNTING	32		SIGHS	25	
GRUNTS	31		CHUCKLE	1	24
LAUGHING	28	46	CHUCKLES	19	
LAUGHS	18		CHUCKLING	4	
GASP	1	43	CRIES	4	24
GASPING	5		CRYING	20	
GASPS	37		PANTING	17	20
YELL	2	37	PANTS	3	
YELLING	28		EXCLAIM	1	18
YELLS	7		EXCLAIMING	14	
SCREAM	4	36	EXCLAIMS	3	15
SCREAMING	25		CHATTER	10	
SCREAMS	7		CHATTERING	5	

興味深いことは、ほとんどの場合、動詞によって原形（または原形 + s）が多いか、原形 + ingの形が多いかに分かれていることである。前者が多いものは、GASP, SIGH, CHECKLE, CHATTERであり、SIGHではing形そのものが出現していない。後者が多いのは、LAUGHING, YELLING, SCREAMING, CRYING, PANTTING, EXCLAMINGである。例外としてGRUNT (S) , GRUNTINGは、32回ずつと半々の割合で出現している。前者の原形型は、比較的短時間の動作や所作を表現しているものであり、後者ING型は時間がある程度続く動作が表現されているからでないかと考えられる。

Zdenekは、サイエンスフィクション4本の映画を調べて、バラ言語字幕として、Gruntが一番多く、Scream, Laugh, Yell, Clamor, Chuckleが多いと報告しているが、今回の調査においても、ほぼ同様の結果となっている ([1])。今回の調査対象となった映画やジャンルがサイエンスフィクションだけでないので、この結果は、バラ言語字幕と表現される動詞で、高頻度のものはジャンルを超えてある程度共通していると考えられる。

5. 3 word2vecを使った類義語との比較

word2vecはニューラルネットワークを使用して単語または語句を実数の数値の並び（ベクトル）に変換するものであり、人工知能を使った自然言語処理の代表的なプログラムとなっている

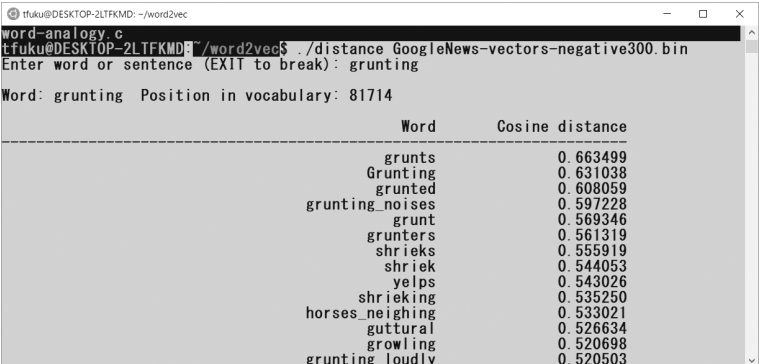
([4] [5] [6])。このプログラムはまず大規模な訓練用テキストデータから語彙を構築し、次にニューラルネットワークを使って単語の分散表現、数百次元のベクトルとして学習するものである。単語がベクトルで表現されるため、ある単語と別の単語がどれだけ類似しているのかを単語ベクトル間のコサイン類似度 (cosine distance) として算出することができる。

実装に当たっては、windows10にてMicrosoft StoreからUbuntu (version 16.04.6 LTS Xenial Xerus) をインストールし、word2vecのパッケージを入れて、これにネット上で入手可能なGoogleNews-vectors-negative300.binの圧縮版を解凍して使用している ([7] [8])。

使用したword2vecは、Google News dataset を入力データとして学習したもので、datasetは1000億語 (100 billion words) から成る ([4])。そこから学習された分散表現は、300万の単語と語句に対して300次元でのベクトル表現となっている。

次に、パラ言語として分類された単語のうち、高頻度の単語 gruntとその派生語について、word2vec プログラムを利用して類義語として算出されるものがどのような語句になるかを調べた。

検索した語は、“grunts”, “grunting”, “grunted” の3単語である。ここでは、コサイン類似度が比較的高い0.50以上のものだけを示している。表6では、三単現のsと、ing形が派生語となっているが、動詞の意味でのgruntをはっきりとさせるため、gruntedと過去形を加えた。



```

word-analogy.c
truku@DESKTOP-2LTFKMD: ~/word2vec$ ./distance GoogleNews-vectors-negative300.bin
Enter word or sentence (EXIT to break): grunting
Word: grunting Position in vocabulary: 81714

```

Word	Cosine distance
grunts	0.663499
Grunting	0.631038
grunted	0.608059
grunting_noises	0.597228
grunt	0.569346
grunTERS	0.561319
shrieks	0.555919
shriek	0.544053
yelps	0.543026
shrieking	0.535250
horses_neighing	0.533021
guttural	0.526634
growling	0.520698
grunting_loudly	0.520503

図2 word2vecでの算出結果画面 (入力語grunting)

表7 gruntsに対するword2vecの結果

grunting	0.663499
grunt	0.611855
yelps	0.580545
guttural	0.547705
shrieks	0.542745
yowls	0.528733
whinnies	0.525461
grunTERS	0.524838
guttural_grunts	0.518494
growls	0.516628
grunting_noises	0.509669
shriek	0.509080
gutteral	0.504628

表8 gruntingに対するword2vecの結果

grunts	0.663498
Grunting	0.631038
grunted	0.608059
grunting_noises	0.597228
grunt	0.569346
grunTERS	0.561319
shrieks	0.555919
shriek	0.544053
yelps	0.543026
shrieking	0.535250
horses_neighing	0.533021
guttural	0.526634
growling	0.520698
grunting_loudly	0.520502
yelping	0.519647
squealing	0.518591

表9 gruntedに対するword2vecの結果

grunting	0.608059	smiled	0.524441
growled	0.606640	quivered	0.523751
muttered	0.586532	twitched	0.521678
giggled	0.577342	scowled	0.516001
hollered	0.571024	hissed	0.513361
mumbled	0.570337	stammered	0.508346
grinned	0.565585	grimaced	0.507700
smirked	0.556664	huffed_puffed	0.506204
murmured	0.551312	glared	0.505815
groaned	0.548331	eyebrows_arching	0.505786
yelped	0.547817	chuckled	0.505196
bellowed	0.544658	laughed	0.503211
cackled	0.536633	yipped	0.502962
barked	0.528173	meowed	0.502490
shrieked	0.527218	jabbered	0.501671
sighed	0.524504	gasped	0.501649

これらの3つの表に共通して出現する語は、yelp, shriek, growlの三個であり、word2vecによる類似度では、gruntに類似した語であると言える。これらは、映画のパラ言語字幕として高頻度で出現する単語とは、ほとんど重複していない。これは、word2vecでの訓練用データであるGoogle Newsが、メディア分野ではあるが、大量の一般的な英語を対象とした学習の結果であるのに対し

て、映画字幕の調査で対象としたのは、数少ない映画での字幕であるため両者の重なりがあるかどうかは、これだけでは判断できない。

word2vecは、学習するデータが違えば、違った分散表現となる。今回の映画字幕が対象としていような口語の表現を集めたデータから学習した結果は、違った分散表現となることが予想されるため、両者の重なりがどれだけあるのかは、実証してみなければ分からないところである。

また、word2vecは単語の意味の表現を表しているものと理解されるが、パラ言語字幕は、あくまで、非発話情報に対する字幕の区分という人為的なものであり、パラ言語字幕に出現する単語間の意味の近さを考慮しての分類ではないため、比較が難しい面がある。

一方で、表9に現れる単語（動詞）は、もし映画中にそれらの動作がなされて、字幕化されたならば、パラ言語字幕と区分されるものが多くあり、word2vecでgruntedと類似度が高いとされる単語群は、パラ言語字幕となる可能性が高いものであると言える。映画字幕のデータがもっと大量にあれば、重なりが増えることが予想される。

6. おわりに

本稿では、5本の英語映画に現れる非発話情報に対する字幕について調査を行い、その結果を分類して分析を行った。分析結果からは、パラ言語字幕や効果音を示す字幕が多く出現することが分かったが、一方で、映画の内容による字幕の区分の多さの違いが明らかとなった。

また、自然言語処理で利用されるword2vecを使って、パラ言語で最も多く表れるGruntについて、どのような類似の語句があるのかを調べ、映画でのパラ言語字幕との相違があるのかを調べたが、両者の重複はほとんどなく、はっきりとした結論がでていない。word2vecが発表されて以降、word2vecよりもより良い分散表現だとされるものも出てきており、今後も、これらの一連の単語の分散表現との比較を続ける（[9]）。

今後は、複数の映画に付与された日本語での難聴者用字幕について調査をし、英語と日本語での非発話情報に対する字幕の類似点、相違点を明らかにする。これを基礎データとして、日本語での非発話情報に対する字幕をどのように付与すべきなのかの指針の確立に役立てるようにする。

7. 謝 辞

本研究はJSPS科研費18K11976（「日本語と英語の字幕の違いに基づいた非発話情報の字幕表示方法の確立に関する研究」）の助成を受けたものです。

参考文献

- [1] S. Zdenek, *Reading Sounds*, The University of Chicago Press, Chicago and London, 2015.
- [2] 福島孝博, 「映画における非発話情報の字幕に関する基本調査 ～同一映画に付与された英語と日本語での難聴者用字幕の比較～」, 電子情報通信学会, 第91回福祉情報工学研究会 WIT2017-17 pp.15-20. August 2017.
- [3] 福島孝博, 「日本語と英語の難聴者用映画字幕における非発話情報としての話者情報, 音楽情報に関する比較調査」, 映像情報メディア学会, 2017年冬季大会, 早稲田大学 December 2017.
- [4] <https://code.google.com/archive/p/word2vec/> (description of word2vec)
- [5] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. “Efficient Estimation of Word Representations in Vector Space”, In Proceedings of Workshop at ICLR (International Conference on Learning Representations), 2013.
- [6] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. “Distributed Representations of Words and Phrases and their Compositionality”, In Proceedings of NIPS (Neural Information Processing Systems), 2013.
- [7] <http://www.ubuntu.com/> (Ubuntu homepage)
- [8] <https://s3.amazonaws.com/dl4j-distribution/GoogleNews-vectors-negative300.bin.gz>
- [9] Jeffery Pennington, Richard Socher, Christopher D. Manning. “GloVe: Global Vectors for Word Representation”, EMNLP (Conference on Empirical Methods in Natural Language Processing) Vol.14. 2014.