

## E-5

# ウェブデータから言語変化を捉える：タガログ語の *sana all* の分析

長屋尚典<sup>1</sup>・林真衣<sup>2</sup>・細羽洗希<sup>2</sup>

**要旨：**本発表では Twitter から得られたデータによってタガログ語の *sana all* の共時的・通時的分析を提示する。本稿の主な主張と発見は以下の通りである。第一に、*sana all* は、好意的に評価される当該の事象に反応して反実仮想と羨望を表現する。第二に、*sana all* は、2017 年ごろから独立用法と前置用法を中心に使用され始め、2019 年 10 月ごろに爆発的に使用が拡散し、用法も多様になった。第三に、*sana all* は、2017 年で既に独立用法・前置用法を持っていた *sana lahat* から生まれたと提案する。最後に、*sana all* がその使用を急速に拡大して以降、その異綴語の種類や用法も増えていったが、綴りの違いが用法の違いに結びついている証拠はなかった。

## 1 はじめに

- 本発表は、コンピューターを介したコミュニケーション (**computer-mediated communication; CMC**) における言語使用と言語変化の一例を分析する：
  - 新語の発生を含む言語変化を促進する側面は長らく注目されている (Shortis 2001)
  - レジスター、異綴語 (同じ語彙素であるが綴りが異なる語) など言語分析の対象としても注目を受けている (Biber & Egbert 2018 など)
  - 今までに得られなかったような言語資源が使用可能になり、新しい言語研究の視野を広げる可能性も持つ (e.g., ウェブコーパス; Twitter API; News on the Web (NOW) corpus)
  - 新型コロナウイルス感染症の感染拡大という社会情勢下においてはその重要性は増している
- 本発表では、タガログ語の最近の流行語である *sana all* [sá:na ʔil] に注目する：
  - この表現はタガログ語の願望小辞 *sana* および英語の *all* からなる複合語である
  - (1)(2) のように、字義通りには *all* 「全ての人 (や物事)」がそうあって欲しいと *sana* 祈念するかのよう表現だが、実際には、**好意的に評価できる事象について、それが話者の現実に当てはまらないので羨ましい・妬ましい**という意味を表現する<sup>3</sup>

(1) Sana+all may=jowa

wish+all EXS=lover

「私に恋人がいればいいのになぁ。」

(2) Sana+all gwapo

wish+all handsome

「私がイケメンだったらいいのになぁ。」

- *sana all* は、2017 年ごろから使用されはじめたと推定できる比較的新しい表現であるものの (後述)、現在では日常会話を含むインフォーマルなレジスターで幅広い年代で使用されている (Pontillas et al. 2020)
- インターネット上では *sana ol* などのいくつかの異綴語が存在するという特徴も持つ
- 本発表ではこの *sana all* の共時的・通時的特徴を Twitter のデータを使用して分析する：
  - 共時的な用法の記述・分析: *sana all* は 2021 年現在どのような用法をもつのか?

<sup>1</sup> 東京大学大学院人文社会系研究科 言語学研究室 nagaya@1.u-tokyo.ac.jp

<sup>2</sup> 東京外国語大学言語文化学部言語文化学科フィリピン語専攻/言語情報コース

<sup>3</sup> 本稿で用いる略号は以下の通り: ADJ-adjective, AV-actor voice, CAUS-causative, EXS-existential, GEN-genitive, IAM-iamitive, LK-linker, LOC-locative, NOM-nominative, NVOL-non-volitional, RDP-reduplicant, RL-realis, SG-singular, <-infix, ”=”-cliticized.

- 通時的な変化の記述・分析: *sana all* はどのように生まれ、どのように変化したのか?
- 綴りの記述・分析: *sana all* の異綴語はどのように生まれ、どのように用いられるのか?
- 本発表の構成は以下の通りである:
  - 第2節: タガログ語と *sana* について背景的知識をまとめる
  - 第3節: 本稿が使用する Twitter データベースについて紹介する
  - 第4節: *sana all* の2021年現在における用法を共時的に分析する
  - 第5節: *sana all* の成立と発展を通時的に分析する
  - 第6節: *sana all* の異綴語のバリエーションについて分析する
  - 第7節: まとめ

## 2 タガログ語と *sana*

- タガログ語はフィリピン共和国マニラ首都圏で話されるオーストロネシア語族の言語である
- 基本語順は VSO または VOS であり、述語が節の先頭に出現し、指示的な名詞句は格助詞または前置詞によって節に導入される
- タガログ語にはさまざまな不変化詞が存在するが、そのひとつに *sana* がある:
  - 節の先頭に現れ、まだ実現してない補文節の内容が実現することを祈念する (例 (3))
  - 第二位置接語として実現しなかった願望 (frustrative) を表現することもある (例 (4))

(3) Sana p(um)asa=ako      sa=exam  
 wish pass(AV)=1SG.NOM LOC=exam  
 「私が試験に合格すればいいなあ。」

(4) Nag-basa=sana=ako      ng=blog    nung=may=WiFi=pa  
 AV.RL-read=wish=1SG.NOM GEN=blog when=EXS=WiFi=still  
 「私は Wi-Fi があるうちにブログを読んだらよかったなあ (現実にはしなかった)。」

- 英語が公用語の一つである事情からタガログ語と英語のコードスイッチングが頻繁に起こり、*sana all* のように両言語の要素からなる表現も頻繁に使用される

## 3 方法

- 本発表では Twitter における *sana all* の用法に絞って分析する:
  - *sana all* のような表現は Twitter のような SNS で (も) 頻繁に用いられる
  - 共時的特徴も通時的変化も異綴語のバリエーションも観察することができる
- 本発表で分析するデータは以下の18のタイムスパンにおける *sana all* (およびその異綴語) である
  - 2017年から2021年までの1月15日、4月15日、7月15日、10月15日(全18日)の12:00~13:00(フィリピン時間)に投稿されたツイート
  - 理由:
    - \* 3ヶ月ごとに同じ日の同じ時間のデータを収集することで *sana all* が成立してから用法が拡大するまでの全体像を描くことができると予想した
    - \* 全用例のアノテーションを実施するため、あまり多くならないようにした

- データ収集方法:
  - R (R Core Team 2020) によって Twitter API を用いて収集した (Barrie & Ho 2021)
  - 検索語: *sana all* およびその異綴語 (*sana ol*, *sana halls*, *sana oil*, *sauna oil*, *china oil*, *naol*)、*sana all* の語源と推定した *sana lahat*
  - リツイートは含まれていない
  - 引用ツイートの引用部分については排除した
  - ユーザー情報等は収集していない
  - Twitter の言語設定については指定していない; 稀に英語のみのツイートやセブアノ語などのツイートがあったが、それは排除した
- この方法によって抽出した 7,487 ツイートをアノテートした:
  - (a) ツイート日時; (b) 返信ツイートか否か; (c) 引用ツイートか否か; (d) 用法
  - 返信ツイートは返信先の記載がある場合を「返信」と判断した
  - 引用ツイートは URL を含むツイートを「引用」と判断した
  - 用法についてはマニュアルでコーディングした (第 4 節)
- 以下で提示する用例は実際のツイートを (個人情報などに配慮して) 適宜修正したものである
- 例文を提示する際には、書記方法はフィリピンで一般的に用いられている表記に改めた
- *sana all* で必ず出現し、その間に他の要素が入らないことから、全体でひとつの複合語と分析し、例文では + を入れて表記している

## 4 *sana all* の共時的分析

- 本節では、2021 年 4 月現在の *sana all* の 6 つの用法を提案し、その意味論的・語用論的特徴について分析する

### 4.1 *sana all* の 6 つの用法

- *sana all* には使用される形態統語論的環境に注目して以下の 6 つの用法を区別することができる
  - (i) **独立 (stand-alone) 用法:** (5)
    - \* *sana all* それ自体で発話を形成する用法で、それ単独で「(現在話題となっている事象について) いいなあ、私もそうならいいのになあ」という感情を吐露する
    - \* 形態統語論的には独立して発話されるが、誰かのツイートに返信したり引用したりするときに頻繁に用いられ (後述)、他のツイートに文脈的に依存している
  - (ii) **前置 (preposed) 用法:** (1)(2)
    - \* *sana all* によって羨望を表現される対象である補文の前に *sana all* が出現する
    - \* 補文で表現された事象について「私もそうであつたらいいのに; うらやましい」という気持ちを表出する
  - (iii) **後置 (postposed) 用法:** (6)
    - \* 前置用法とは逆に羨望の対象となる補文の後ろに *sana all* が出現する
    - \* 前置用法との意味論的違いは今のところ観察されていない
  - (iv) **述語 (predicate) 用法:** (7)
    - \* *sana all* 全体で「*sana all* という気持ちである; 羨ましい」という述語を形成する

\* 「羨ましい」という感情をいやく主体が主格名詞句として表現され後続する

– (v) 語基 (base) 用法: (8)

\* *sana all* が述語派生の語基として使用される

\* *ma-pa-sana+all* 「思わず *sana all* と言ってしまう」、*mag-sana+all* 「*sana all* と発言する」、*taga-sana+all* 「*sana all* と言っただけで実際に得られたことがない」などの派生語が存在する

– (vi) その他 (other): 引用表現の中に現れる場合や名詞として用いられる場合がある

(5) Sana+all 🥰🥰

wish+all EMOJI

「羨ましいなあ。」(返信ツイート)

(6) Best couple sana+all

best couple wish+all

「ベストカップル、羨ましいなあ。」

(7) Sana+all=na=lang=ako

wish+all=IAM=only=1SG.NOM

「私は *sana all* っていう気持ちだ。」

(8) Ma-pa~pa-sana+all=ka=na=lang=talaga

NVOL-RDP~CAUS-wish+all=2SG.NOM=IAM=only=really

「あなたは本当に *sana all* と言うしかないだろう。」

## 4.2 *sana all* の意味論

- *sana all* の意味には、好意的に評価される当該の事象について「話者の現実ではそうではない」という反実仮定の側面と「羨ましい、自分も手に入れたい」という羨望の二つの大きな側面がある:
  - ただし、どちらの側面が前景化されるかには違いがあり、同じ *sana all* でも単なる嫉妬から、(消極的な) 賛辞、そして冗談までさまざまなニュアンスがある
  - タガログ語母語話者によれば、どのニュアンスになるかは使われる文脈やイントネーション、使う人の人間性、お互いの関係性に依りて異なる
  - (Twitter のデータではそのようなニュアンスを正確に判断することは難しい)
- *sana all* を使って羨む事象は恋愛や美容、芸能、(韓流) アイドルに関する内容に大きく偏るという特徴が観察される:
  - これは Twitter というメディアやそのユーザーの年齢層の反映に過ぎないかもしれない
- *sana all* の羨望の意味は単なる希求表現の *sana* と比較することでよくわかる:
  - (9) のように *sana* 単独では単に当該の事象を中立的に希求するのみであり、特に誰かを羨ましく思っているという意味はない
  - 一方で、(10) のように *sana all* であれば、自分がセメスター途中の休みを希求する一方で、実際にセメスター途中の休みを迎えた他人の存在が含意され、その結果、その他人を羨む話者の嫉妬が表出している

(9) Sana sembreak=na

wish semester.break=IAM

「もうセメスター途中の休みだったらなあ。」

(10) Sana+all sembreak=na

wish+all semester.break=IAM

「私もセメスター途中の休みだったらなあ。」

## 4.3 *sana all* の CMC 上の特徴

- *sana all* は返信ツイートまたは引用ツイートで用いられやすいという特徴がある

表1 *sana all* 独立用法の返信・引用の分布

	返信	↯返信	合計
引用	17 (337.1)	618 (297.8)	635
↯引用	960 (639.8)	245 (565.2)	1205
合計	977	863	1840

表2 *sana all* 前置用法の返信・引用の分布

	返信	↯返信	合計
引用	12 (181.0)	404 (235.0)	416
↯引用	780 (611.0)	624 (793.0)	1404
合計	792	1028	1820

- 特に独立用法でその傾向が顕著である; 表1は、独立用法の *sana all* を含むツイート 1,840 件に関して、返信ツイートであるか、引用ツイートであるかを分析したものである
- 独立用法の *sana all* がよく使われる状況としては (引用ではない) 返信ツイートと (返信ではない) 引用ツイートがあり、返信も引用もしない (ふつうの) ツイートではあまり使用されない
- 前置用法の場合でも同様の傾向が観察されるが、返信でもなく引用もしないツイートの数が独立用法に比べて増えている (表2)
- 比較として 2021 年 5 月 14 日における *lahat* 「全て」を含むツイート 2,000 件を調べたところ、最も多かったのは返信も引用もしない (ふつうの) ツイートで、1,258 件 (62.9%) であった
- このことは *sana all* が単なる羨望の表明ではなく、CMC における反応としての反実仮想・羨望を表現しているという特徴があることを示している

## 5 *sana all* の通時的分析

- 本節では第4節で示した *sana all* の用法がどのように通時的に形成されてきたかを論じる

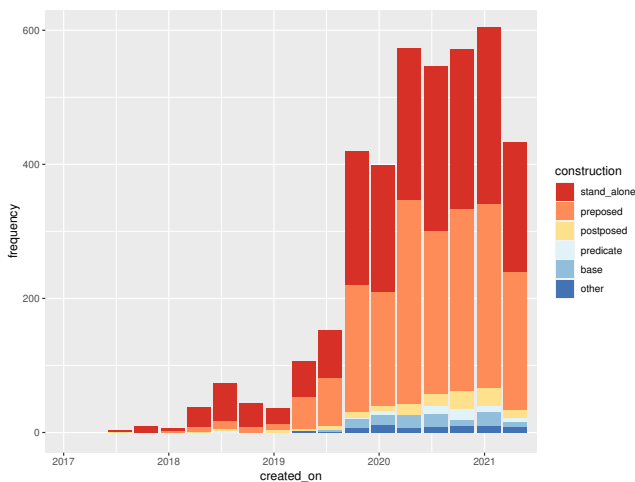


図1 *sana all* の用法の変遷

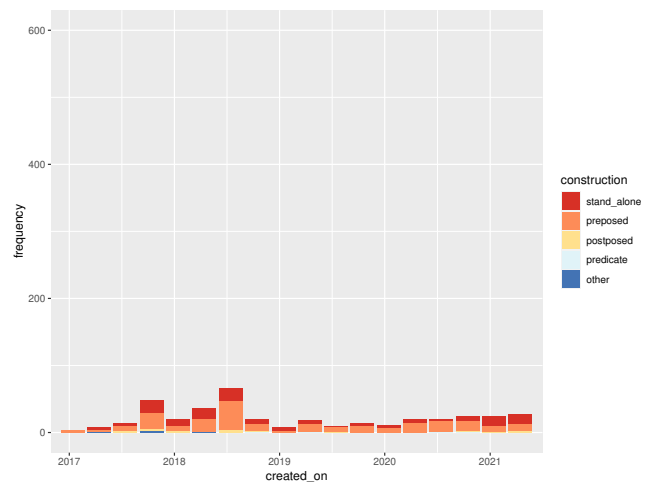


図2 *sana lahat* の用法の変遷

### 5.1 *sana all* の成立と変遷

- 2017 年から 2021 年までの *sana all* の各用法の頻度をまとめた (表3・図1)
- *sana all* の成立:
  - *sana all* は少なくともツイッター上では 2017 年 7 月あたりから徐々に使用されはじめ、2019 年 10 月ごろに急速に拡散 (go viral) し、その後、安定して使用されている
  - このデータからは日常会話でこの表現が使用され始めたのか、それとも Twitter が先行しているのかはわからない

- *sana all* の変遷:
  - 2018 年までは独立用法が支配的であり、その後、前置用法が増加した
  - 2021 年現在では独立用法と前置用法の例がほとんどをしめる
  - 2019 年 10 月に急速に拡散したあたりから用法も多様化してほぼ現在の用法が出そろった
- *sana all* の成立自体は漸次的 (gradual) な変化であるが、用法の拡大は急速 (abrupt) な変化とみることができるかもしれない
- この分析は、2021 年 4 月までの *sana all* の分析であり、*sana all* の変化は現在も進行中である

## 5.2 *sana all* の起源

表 3 *sana all* の用法の変遷

ツイート日	独立	前置	後置	述語	語基	その他	合計
2017-01-15	0	0	0	0	0	0	0
2017-04-15	0	0	0	0	0	0	0
2017-07-15	2	0	1	0	0	0	3
2017-10-15	9	0	0	0	0	0	9
2018-01-15	4	2	0	0	0	0	6
2018-04-15	29	7	1	0	0	0	37
2018-07-15	57	12	3	2	0	0	74
2018-10-15	35	8	0	0	0	0	43
2019-01-15	23	10	3	0	0	0	36
2019-04-15	54	47	3	0	0	2	106
2019-07-15	71	73	6	0	2	1	153
2019-10-15	200	188	10	2	12	7	419
2020-01-15	190	170	8	5	15	11	399
2020-04-15	226	305	16	0	19	7	573
2020-07-15	246	244	18	12	19	8	547
2020-10-15	238	272	26	16	10	9	571
2021-01-15	263	275	27	8	22	9	604
2021-04-15	193	207	12	6	7	8	433
合計	1840	1820	134	51	106	62	4013

- 本研究は *sana all* の起源として *sana lahat* を提案する (*lahat* 「すべて」)
- *sana lahat* は既に 2017 年の時点で *sana all* と共通する独立用法も前置用法をもち、「全ての人(事物)が...でありますように」という祈念を表現していた (11)(12) (図 2):

– *sana* のあとに *all* が後続

する構文は *sana* の補文内において *lahat* を含む主語が倒置されることで生じる

- たとえば、(12) は ***Sana magaling sa math (ang) lahat*** という語順の方がより規範的である
- *sana lahat* には *sana all* に見られるような羨望はあまり観察されず、話題も恋愛や美容に関することに限られない
- *sana lahat* の *lahat* が英語の *all* に置き換わることで、特殊な話題について羨望の意味で使用される表現 *sana all* が成立したのではないか<sup>4</sup>

(11) *Sana+lahat* 🙏🙏🙏 [URL]

wish+all EMOJI

「みんな (URL のようであれば) いいなあ。」

(12) *Sana+lahat ma-galing sa=math*

wish+all ADJ-good LOC=math

「みんな数学が得意であればいいなあ。」

- *sana all* の起源 (仮説 2):
  - ***Sana [all people can be who and where they want to be].*** のような *sana* のあとに英語の文が続くような構造から、前置用法の *sana all* が再分析されたという仮説も考えられる
  - この場合、*sana all* の用法として初期に (前置用法ではなく) 独立用法が支配的であった理由が説明できない

<sup>4</sup> *sana all* が返信または引用とともに用いられることが多いのに対して、*sana lahat* ではこの傾向は見られず、返信でもなければ引用もしていないふつうのツイートが一番多かった (388 例中 161 例 (41.5%))。

## 6 *Sana all* とその異綴語

- *sana all* には音声的に類似する単語を使用した (広い意味での) 異綴語が存在する
- *sana all* の異綴語の成立と変遷 (図 3):
  - 異綴語は時間が経つにつれてどんどん増えている
  - *sana ol* や *naol* は *sana all* を追うように頻度と用法を増やしている
  - 一方で *sana oil* のように早い段階で使われなくなったものも存在する
- 綴りの違いが意味や用法の違いに反映していることを示唆するデータは得られなかった (表 4)

## 7 おわりに

- 本発表は Twitter から得られたデータによって *sana all* の分析を行った:
  - 好意的に評価される当該の事象に**反応して反実仮想と羨望**を表現する
  - 2017 年ごろから独立用法と前置用法を中心に使用され始め、2019 年 10 月ごろに急速に拡散し、用法も多様になった
  - 2017 年で既に独立用法・前置用法を持っていた *sana lahat* から生まれたと仮定できる
  - 使用が急速に拡大するにつれ、その異綴語の種類も用法も増えたが、綴りの違いが用法の違いに結びついている証拠はなかった
- 方法論としては、Twitter データの利用や、CMC の特徴 (返信・引用) を活かした分析を提案した

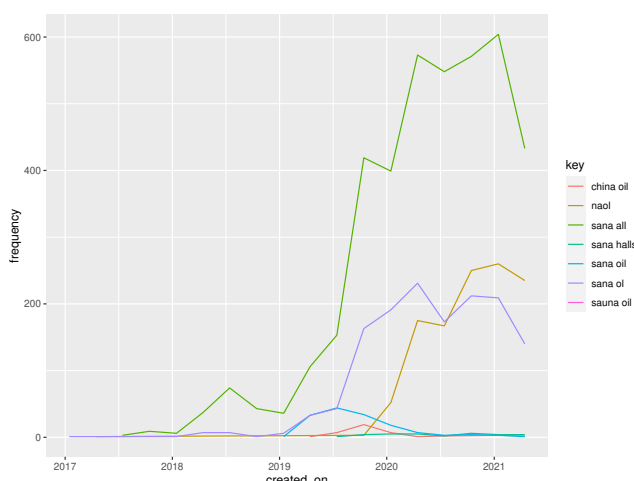


図 3 *sana all* とその異綴語の頻度の変遷

**謝辞:** 本稿に関する内容については以下の方から貴重な意見および情報をいただいた: Violy Bolaño、Justine Joyce Dabalos、John Oliver Monghit、大西貴也、島健太、鈴木唯、谷川みずき、内原洋人、山本恭裕、吉田樹生 (敬称略)。言うまでもなく本稿に残るいかなる誤りも著者の責任である。なお、本研究は JSPS 科研費 JP18H03658 (PI: 投野由紀夫)・JP21K00522 (PI: 長屋尚典) の助成を受けたものである。

表 4 2020 年以降の *sana all* とその異綴語の用法

	<i>sana all</i>		<i>sana ol</i>		<i>naol</i>	
独立	1356	(43.4%)	526	(45.5%)	527	(46.3%)
前置	1473	(47.1%)	559	(48.4%)	578	(50.7%)
後置	107	(3.4%)	30	(2.6%)	25	(2.2%)
述語	47	(1.5%)	12	(1.0%)	7	(0.6%)
語基	92	(2.9%)	15	(1.3%)	0	(0%)
その他	52	(1.7%)	14	(1.2%)	2	(0.2%)
合計	3127	(100.0%)	1156	(100.0%)	1139	(100%)

## 参考文献

- Barrie, Christopher & Justin Chun-ting Ho. 2021. *academictwitterR: an R package to access the Twitter Academic Research Product Track v2 API endpoint*. R package version .0.0.9000. <https://github.com/cjbarrie/academictwitterR>. doi:10.5281/zenodo.4714637
- Biber, Douglas & Jesse Egbert. 2018. *Register Variation Online*. Cambridge: Cambridge University Press.
- Pontillas, Marlon, Elaine Mae Tormes, Zairah Shane Barcelá, Ma Christina Lorzano, Lian Shane Naparato & Jessa Folloso. 2020. *Sana all: Filipino teenagers' indirectness of today's time*. DLSU Research Congress 2020, De La Salle University, Manila, Philippines, June 17-19, 2020.
- R Core Team. 2020. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Shortis, Tim. 2001. *The Language of ICT: Information and Communication Technology*. London: Routledge.