Mining for Goals: Text Mining as a Method to Characterize Goals in the Wild

Caroline A. Hamilton, Hannah Moshontz, Erin K. Davisson, and Rick H. Hoyle

Duke University

Background

Goal setting theory asserts that the *type* of goal one sets matters—a specific, difficult goal is more successful than an easy goal, a vague goal, or no goal at all. Extant research on goal properties and goal setting has largely focused on the effects of goal properties on *performance,* particularly in the context of *work-related* tasks. Relatively less is known about how these effects generalize to spontaneously set personal goals, such as New Year's resolutions.

***Research aim*:** To use **self-report data** and **text analysis** to 1) **characterize** New Year's resolutions, and 2) **identify resolution content** associated with subjectively reported success *on an exploratory basis.*

Method

Design: Longitudinal study with four waves; unit of analysis: resolution
Subjects: 413 US Mturk workers (52% female, 81.8% Caucasian, M_{age} = 36) who set at least one resolution for 2016

T1 January 2016 N = 1,090 resolutions

- Detailed information about up to 5 previously set resolutions

T4 January 2017 N = 548, 50%

- Status and progress of each resolution
- Detailed ratings of success

**Note: T2 (April 2016) & T3 (July 2016) were part of a separate study.

1090 resolutions

Clean text → **Remove function words** → **Tokenize** → **Word frequencies**

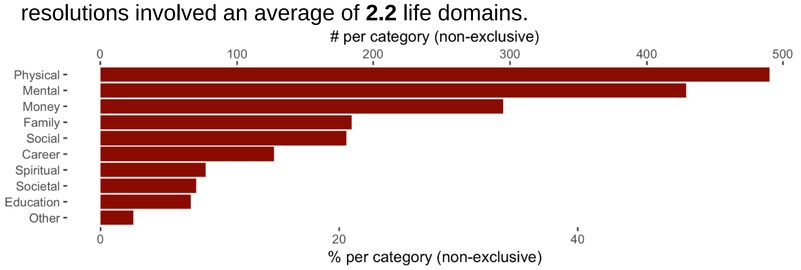
Group successful vs. unsuccessful

Remove function words → **Tokenize** → **Word frequencies**

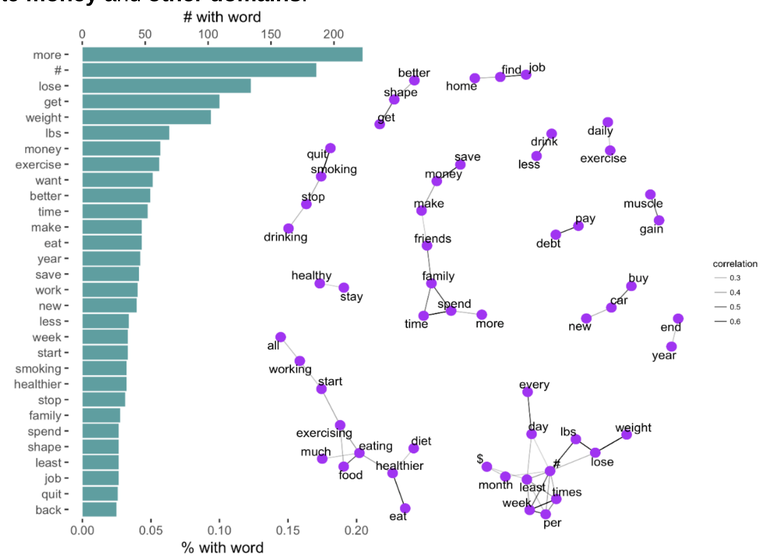
Results

What resolutions did people set?

There was considerable **category crossover**—participants reported that resolutions involved an average of **2.2** life domains.



Most prevalent resolution words related to **physical health**, and fewer related to **money** and **other domains**.

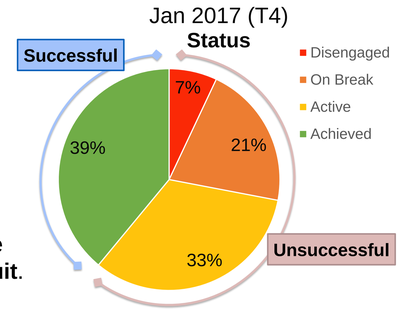


Figs. 1 & 2. (1) Most prevalent word frequencies across resolutions and (2) correlation network of co-occurring words. All displayed words in correlation network were used at least 10 times across the resolution text and denote correlations with Φ > 0.2.

Were resolutions successful?

Of the 548 resolutions for which participants reported the status at T4, **215 (39%) were successful** and **333 (61%) were unsuccessful**.

Most failed resolutions were unsuccessful because participants reported being **"on break"** or **still engaged in resolution pursuit** by the end of the year. **Few were actively quit**.



Successful resolutions were similar to unsuccessful resolutions

Fig 3. Comparison cloud of successful vs. unsuccessful resolutions. Larger words denote larger differences in relative frequency.

Successful resolutions contained more **relative terms** (e.g., "more", "improve", "better") while unsuccessful resolutions contained more **specific terms** (e.g., numbers, "lbs", "stop").

There were **no significant differences in word frequencies**, however (Table 1).

word	proportion of successful resolutions (N = 215)	proportion of unsuccessful resolutions (N = 333)	proportion of all resolutions (N = 548)	*p*
more	.23	.18	.20	1.00
#	.12	.23	.19	.18
exercise	.09	.06	.07	1.00
weight	.08	.08	.08	1.00
lose	.07	.15	.12	1.00
better	.07	.04	.05	1.00
make	.07	.05	.06	1.00
work	.07	.04	.05	1.00
money	.05	.05	.05	1.00
save	.05	.04	.04	1.00
week	.04	.04	.04	1.00
lbs	.02	.09	.07	.17
stop	.02	.05	.04	1.00

Table 1. Relative frequencies of some of the most prevalent words used in resolutions. There were no significant differences in word frequencies.

Phrasing of weight loss resolutions was not related to success

There was **not a significant difference** in the success of **vague** ("lose weight") vs. **specific** ("lose X lbs") weight loss resolutions.

	"lose weight"	"lose X lbs"	total
successful	12	5	17
unsuccessful	23	29	52
total	35	34	69

Table 2. Frequency of successful vs. unsuccessful weight loss resolutions using "lose weight" vs . "lose X lbs" phrasing; $X^2(1, N = 69) = 2.58$, $p = 0.11$.

Weight loss resolution phrasing was **not related to participants' continuous ratings of success**, either.

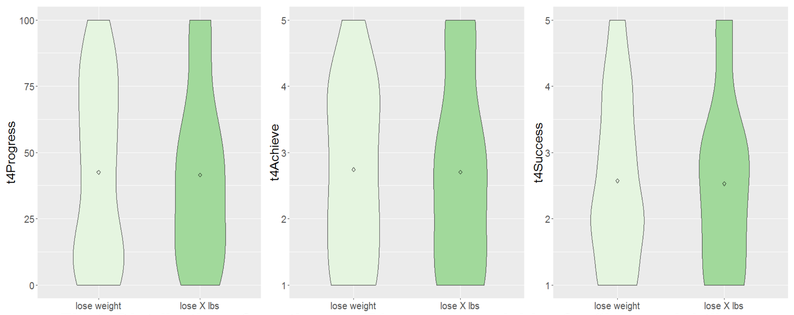


Fig 4. Distribution of continuous T4 success variables for "lose weight" vs. "lose X lbs" resolutions; p > 0.89 for all variables.

Conclusions

These analyses did not offer evidence to suggest that the resolution itself matters to success—successful vs. unsuccessful resolutions were more similar than different with regards to resolution content. This pattern was consistent when examining how weight loss resolutions were phrased. In everyday goal pursuits with spontaneously-set goals**, goal content may matter less than motivational and individual difference variables**.

Limitations: Resolution length, sample size, nesting, synonyms, polysemy
***Future directions*:** Re-running study for 2018 resolutions with a greater sample size (N = 799, 1516 resolutions) and targeted free response questions (~50 words per response).

Full materials and more info available at https://osf.io/y4rt2/ Contact: cah78@duke.edu | ham24@duke.edu