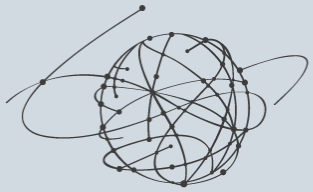
The De-jargonizer: Technology for assessing vocabulary when communicating with expert and non-expert audiences

Tzipora Rakedzon[1,2], Elad Segev[3], Noam Chapnik[3], Roy Yosef[3] Ayelet Baram-Tsabari[1]

1 Faculty of Education in Science and Technology, Technion- Israel Institute of Technology, Haifa, Israel
2 Department of Humanities and Arts, Technion- Israel Institute of Technology, Haifa, Israel
3 Department of Applied Mathematics, Holon Institute of Technology, Holon, Israel

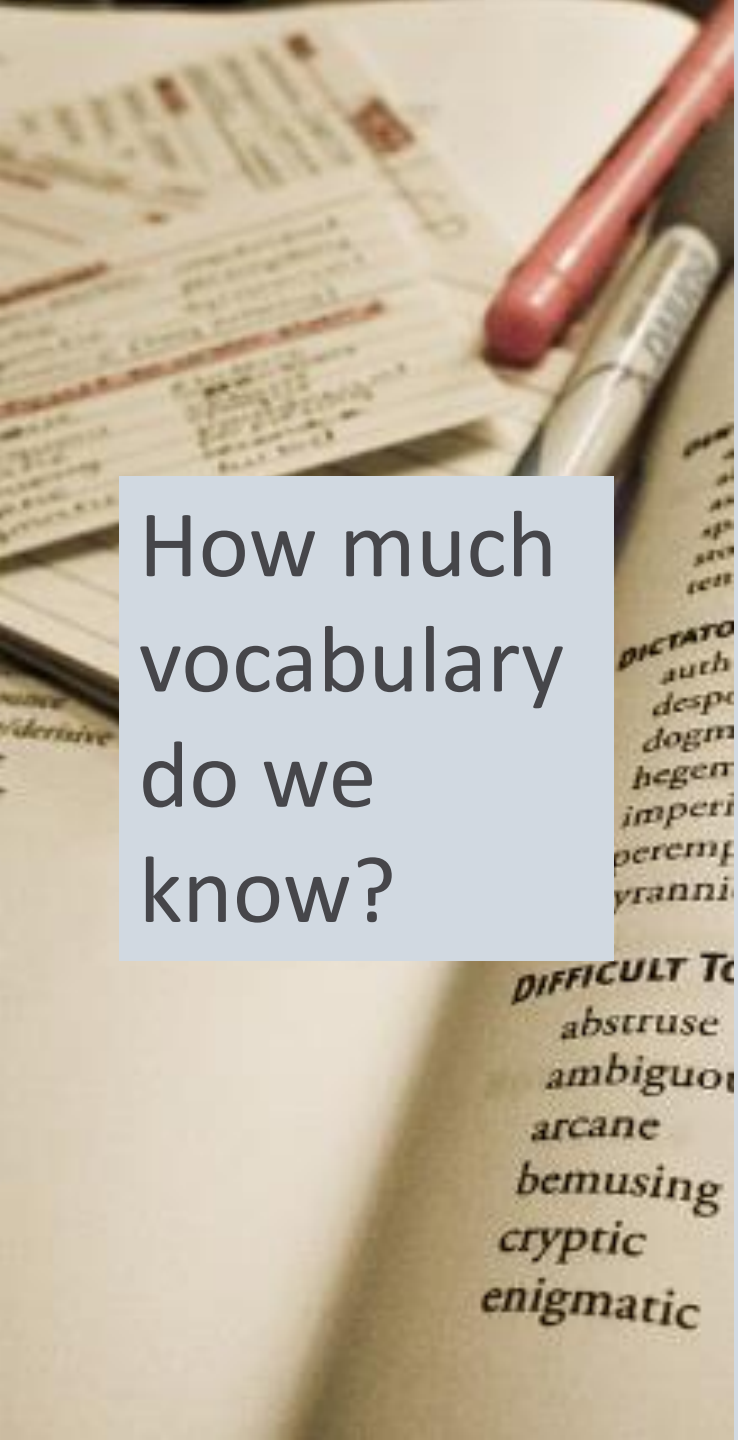
Teaching and Assessing STEM Graduate Students' Scientific Writing for Academia and for the Public



LINKS Learning In a NetworKed Society

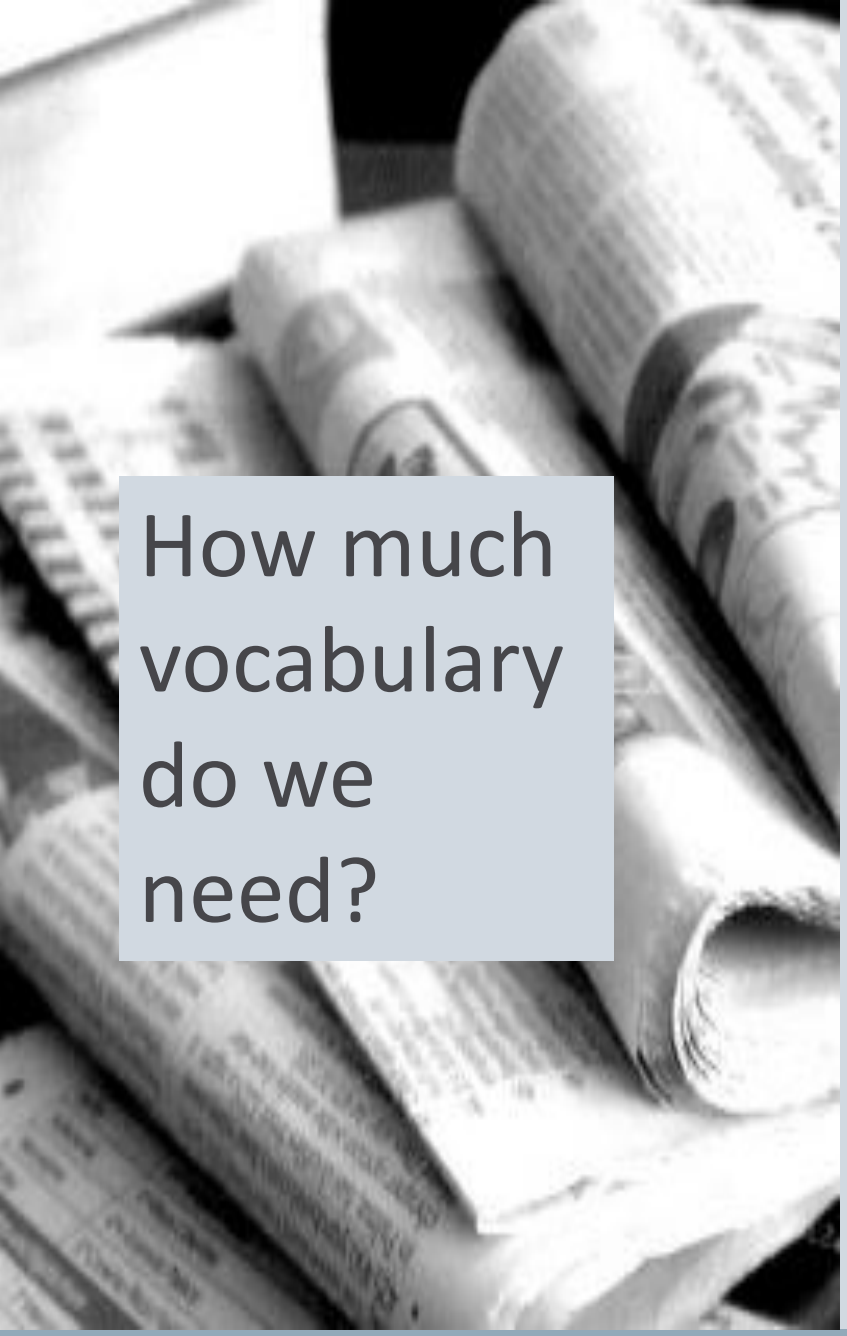
Vocabulary?

- The problem – we need vocabulary for communication → spoken and written
- How much vocabulary –
  - do we know?
  - do we need?



How much vocabulary do we know?

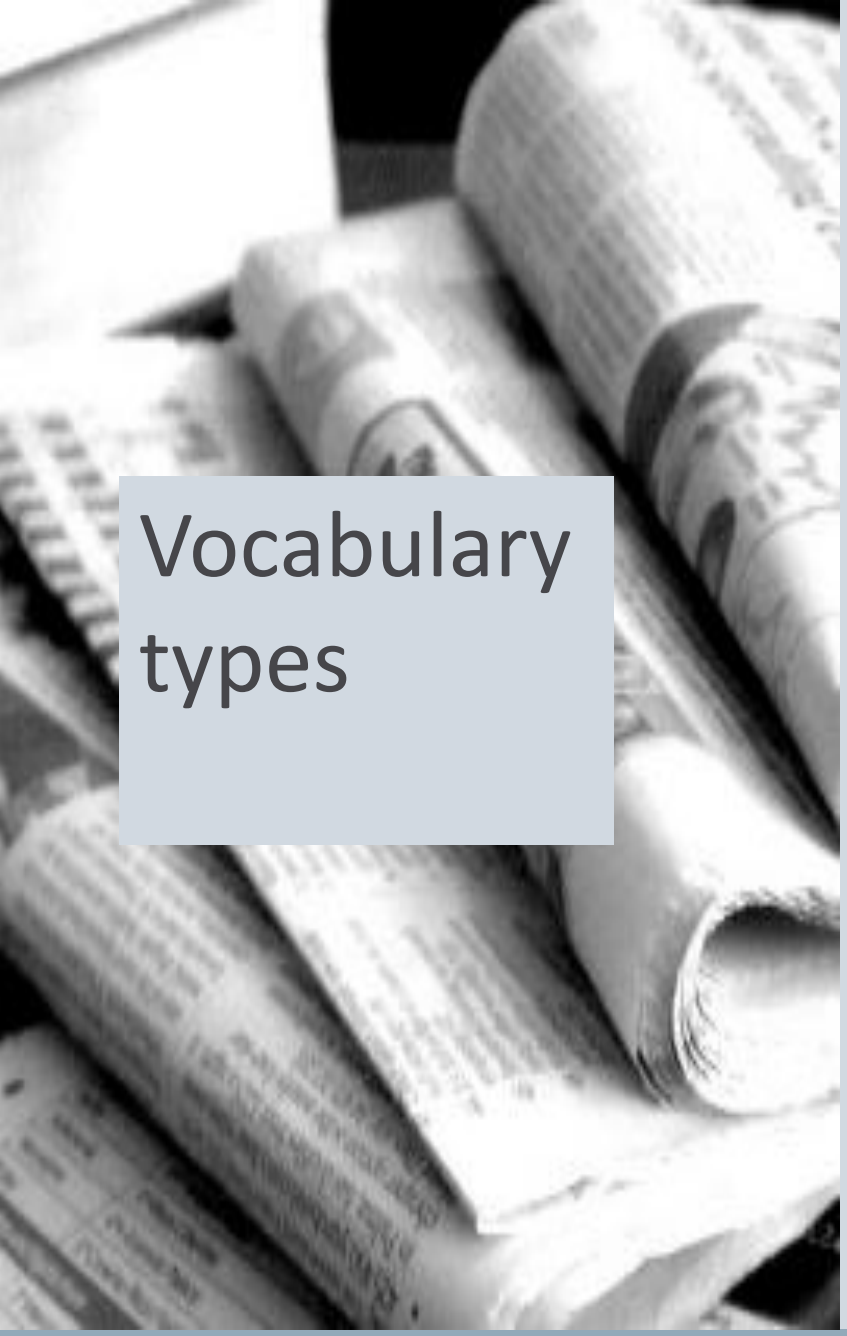
Speakers of	Vocabulary
Native: Typical English-speaking American	42,000 words by age 20, about 1000 word families per year (Brysbaert, Stevens, Mandera & Keuleers, 2016; D'Anna, Zechmeister, & Hall, 1991)
Nonnative: highly educated	Around 8,000 to 9,000 word-families (Nation, 2006)
Nonnative: most foreign test-takers	4,500 words; median - 7,826 (The economist, 2011; online survey ('test your vocab') of over two million worldwide participants)



How much vocabulary do we need?

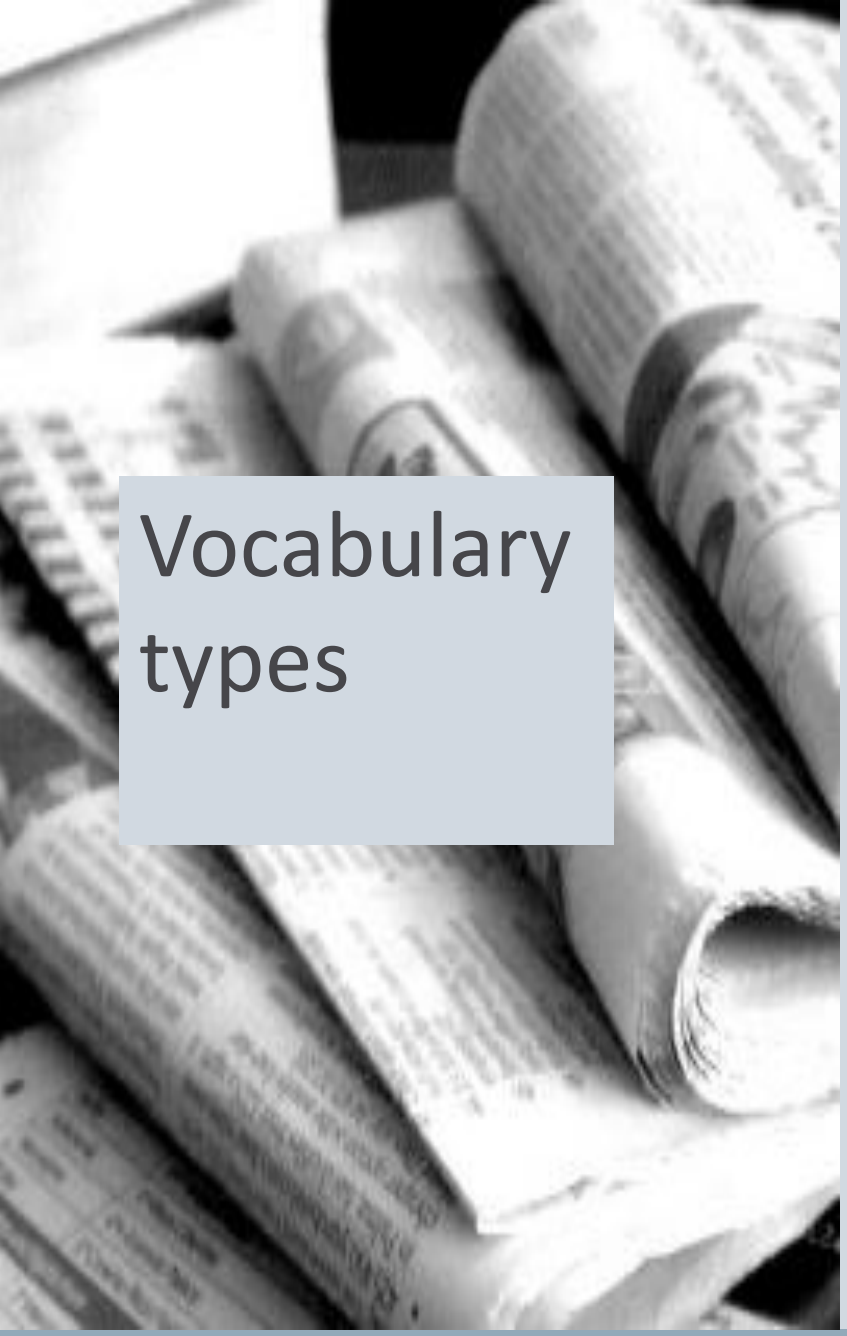
- 8,000–9,000 word-families for written text (novels, newspapers)
- 6,000–7,000 families for dealing with spoken text
- To understand written and spoken language - 98% ideal

(Nation, 2006)



Vocabulary types

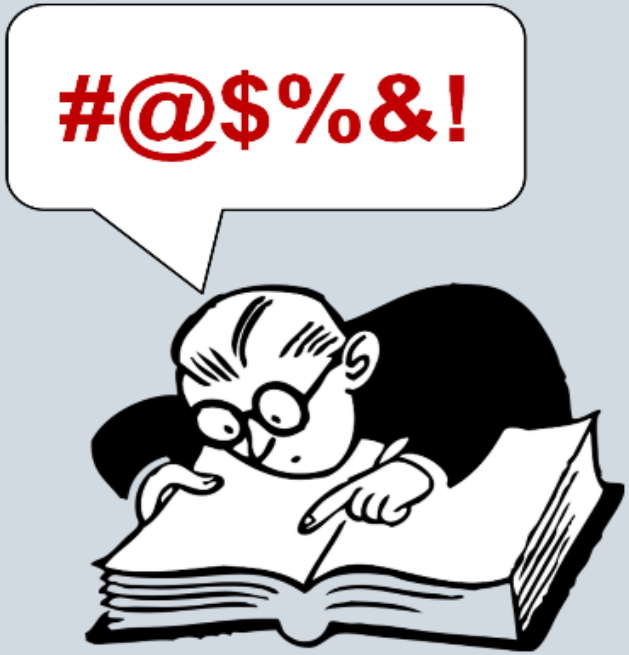
High-frequency words	first 2000-3000 most frequent families
Mid-frequency	word families 4000 to 9000
Low-frequency words	word families 10-25th thousand



Vocabulary types

Academic	many mid-frequency, across disciplines
Content specific /technical	medicine, agriculture, chemistry, applied linguistics, engineering, law, computer science

The problem……



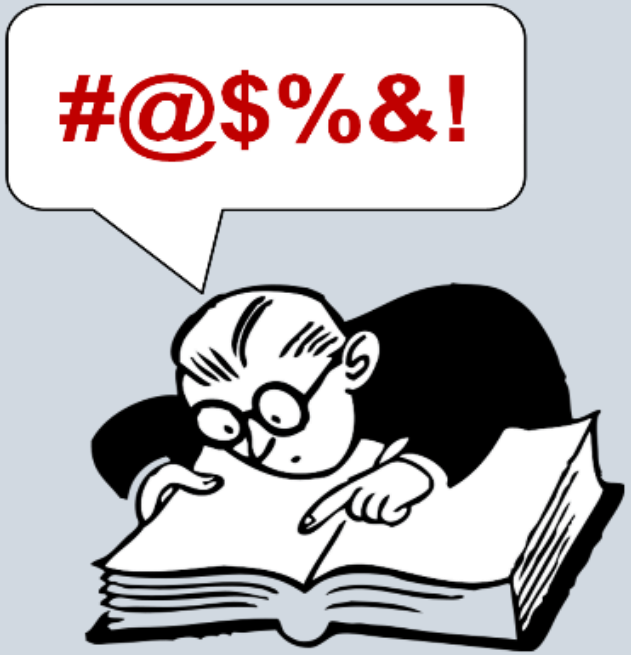
- Readers need to be familiar with 98% of the words in a text for comprehension (Hu & Nation, 2000; Laufer, 1998)

- Vocabulary needs to be audience-appropriate

e.g. Academic and scientific texts have 5-22% jargon (Nation, 2001, Hyland & Tse, 2007) → too much **jargon** for non-experts

The problem……

Limited existing programs for assessing vocabulary/jargon:



- VPcompleat is suitable for researchers in language/linguistics https://www.lextutor.ca/vp/comp/

- Baram-Tsabari & Lewenstien (2013) and Sharon & Baram-Tsabari (2014)

→ did not produce a consistent or usable program

§

VPcompleat example

Freq. Level	Families (%)	Types (%)	Tokens (%)	Cumul. token %
K-1 Words :	54 (50.47)	61 (48.41)	106 (56.38)	56.38
K-2 Words :	21 (19.63)	22 (17.46)	27 (14.36)	70.74
K-3 Words :	21 (19.63)	22 (17.46)	26 (13.83)	84.57
K-4 Words :	1 (0.93)	1 (0.79)	2 (1.06)	85.63
K-5 Words :	2 (1.87)	2 (1.59)	2 (1.06)	86.69
K-6 Words :	2 (1.87)	2 (1.59)	2 (1.06)	87.75
K-7 Words :				
K-8 Words :	2 (1.87)	2 (1.59)	2 (1.06)	88.81
K-9 Words :	1 (0.93)	1 (0.79)	1 (0.53)	89.34
K-10 Words :				
K-11 Words :	1 (0.93)	1 (0.79)	5 (2.66)	92.00
K-12 Words :				
K-13 Words :				
K-14 Words :	1 (0.93)	1 (0.79)	1 (0.53)	92.53
K-15 Words :				
K-16 Words :	1 (0.93)	1 (0.79)	1 (0.53)	93.06
K-17 Words :				
K-18 Words :				
K-19 Words :				

VPcompleat example

magnetic skyrmions are nanometer scaled spin configurations which carry a nontrivial topological charge as a result they cannot be annihilated by any continuous deformation and are therefore very robust despite their small size the topological charge also gives rise to the topological hall effect which presents the means for simple electrical detection of skyrmions their unique properties make them leading candidates for applications in ultra dense magnetic storage applications skyrmions are stabilized by the chiral dzhiloshinski moriha interaction which favors perpendicular spin configurations recently a platform which hosts stable skyrmions at room temperature has been realized in thin multilayer films of irfe co pt where and denote the thickness of the fe and co layers respectively in this project we conduct a thorough quantitative study in order to relate the topological hall signal we obtain from electrical transport measurements with the actual skyrmion density we obtain from magnetic force microscopy images by comparing these two methods we can test theories describing the topological hall effect and achieve a better understanding of this important physical phenomena this understanding is a crucial milestone for future technological applications of magnetic skyrmions

The solution?

Training:
students and professionals in awareness of audience

- for communicating with the non-specialist public [doctor-patient; lawyer-client; scientist- public]
- teachers can manipulate the audience for whom the students write (Chen, 2013)
- E.g. fifth grade students vary their vocabulary by genre; persuasive text more diverse than informative text (Olinghouse & Wilson ,2013)

The solution?

- *Technology:*
  - to help assess vocabulary to aid in adjusting texts

e.g. Academic articles to popular science

The technology

An objective, computerized tool for assessing the use of vocabulary and, specifically, jargon….

- Aiding teachers, scientists and students to easily identify jargon
- Identifying problematic vocabulary and adjust texts accordingly

The De-jargonizer

http://scienceandpublic.com/

identifies vocabulary according to **three** levels:

Word classification	Examples
High frequency - **black**	Pressure, current
Mid frequency – **yellow**	Genetic, protein
Jargon - **red**	Phylogenetics, ions

De-Jargonizer Multiple Texts Hebrew About Instructions Examples Development Developers How to cite Contact Us

De-Jargonizer

How accessible is your work? Paste your article or upload a file to analyze the amount of jargon in your writing.

Time Period

2012 - 2015

Article

Choose File No file chosen

You can also insert the text manually:

Start

Result

Text example: Professional journal abstract

Ants cushion applied stress by active **rearrangements**

Fire **ants**, **Solenopsis invicta**, link their bodies together to form **waterproof rafts**, which in turn **drip**, spread, and **coagulate**, **demonstrating** properties of an active material that can change state from a **liquid** to a solid. This soft-matter phase transition is important when the **raft interacts** with environmental forces such as rain-**drops** and **crashing** waves. We study this active behavior through plate-on-plate **rheology** on the **ants**, **extracting** the active **components** by comparison with the **rheological** behavior of a collection of dead **ants**. In controlled **shear** tests, both and live and dead **ants** show properties of a non-**Newtonian fluid**, specifically, **shear**-**thinning** behavior. In **oscillatory** tests, live **ants exhibit** a rare behavior in which their storage **modulus** (G') and loss **modulus** (**G''**) have approximately the same value over three orders **magnitudes** of **frequency** and two orders of **magnitude** of strain, **indicating** the **ants** are neither **fluid** nor solid. In comparison, dead **ants** are more solid-like, with a storage **modulus** twice as large as their loss **modulus**. This striking active behavior **arises** from **rearrangement** of their bodies and storage and **dissipation** of energy with the **ants' muscles**.

Common:	76%, 149
Mid-Frequency:	15%, 30
Rare:	9%, 18
Suitability for general audience score:	83
Number Of Words:	197

Bulletin of the American Physical Society 66th Annual Meeting of the APS Division of Fluid Dynamics Volume 58, Number 18

Text example: New York Times coverage

Ants That Can Flow Like a **Fluid**, or Move Like a Solid

Fire **ants**, those **stinging pests** that are all too familiar to **Southerners**, have **gotten** plenty of attention from scientists — but not from **physicists**.

Usually the issue is how to stop the spread of the most **problematic** of the different species, or figure out why the **sting** is so painful. But scientists at the Georgia Institute of Technology, including **Zhongyang Liu** and David **Hu**, were interested in the ways that a mass — or you might say, a mess — of fire **ants** can act like a **fluid** or a solid, depending on the situation. It's the first time this **duality** had been observed in a group of living things.

In a **presentation** at a meeting of the American Physical Society last month the researchers showed video of the **ants pouring** out of a **funnel**, like some thick and **wriggling syrup**, and also **springing** back into a rough ball shape after being pressed down.

These images **illustrated** the findings of the more technical research, done with **rheometers** to measure the precise **viscosity** and **elasticity** of balls of **ants** under stress. The researchers found that in different situations the **ants behaved** differently.

Common:	85%, 171
Mid-Frequency:	12%, 24
Rare:	3%, 6
Suitability for general audience score:	91
Number Of Words:	201

http://www.nytimes.com/2013/12/17/science/ants-that-can-flow-like-a-fluid-or-move-like-a-solid.html?ref=science

Student's work: academic abstract

Magnetic skyrmions are nanometer scaled spin configurations which carry a nontrivial topological charge. As a result, they cannot be annihilated by any continuous deformation, and are therefore very robust despite their small size. The topological charge also gives rise to the topological Hall effect, which presents the means for simple electrical detection of skyrmions. Their unique properties make them leading candidates for applications in ultra-dense magnetic storage applications. Skyrmions are stabilized by the chiral Dzhiloshinski- Moriha interaction which favors perpendicular spin configurations. Recently, a platform which hosts stable skyrmions at room temperature has been realized in thin multilayer films of Ir\Fe[x]\Co[y]\Pt, where [x] and [y] denote the thickness of the Fe and Co layers respectively. In this project we conduct a thorough quantitative study in order to relate the topological Hall signal we obtain from electrical transport measurements, with the actual skyrmion density we obtain from magnetic force microscopy images. By comparing these two methods we can test theories describing the topological Hall effect, and achieve a better understanding of this important physical phenomena. This understanding is a crucial milestone for future technological applications of magnetic skyrmions.

Common:	72%, 139
Mid-Frequency:	15%, 29
Rare:	13%, 25
Suitability for general audience score:	80
Number Of Words:	193

Student's work: same topic, for non experts

A magnet can be thought of as an array of tiny magnets called spins, each pointing in their own direction. In a standard magnet, like the ones we see around us in everyday uses, the spins pay an energy penalty for not being aligned, and thus generate a strong magnetic field. Although the majority of spins point in the same direction, sometimes there are domains in which the spins point in the oppositely. Only the spins at the edge of the domain pay a penalty for not aligning with their neighbors, therefore small domains in which the energy penalty from the edges is bigger than the energy reduction at the bulk become unstable and collapse. Surprisingly, some special materials have an energy penalty for parallel spins. Those materials can host very small magnetic domains called "skyrmions". The size of a skyrmion is typically a few nan-meters (10^-9 m), and they are actually very stable. When electrons pass through a skyrmion, they are deflected sideways and give rise to a measurable voltage. This effect is called the "topological Hall effect", and it offers a simple method for skyrmion detection. Skyrmions can therefore be used for very dense data storage. In our project, we obtain an image of the magnetic field the skyrmions generate using a device called a "magnetic force microscope", and compare the image to voltage measurements. This way we learn about the topological Hall effect, which is an important milestone for technological applications of skyrmions.

Common:	81%, 200
Mid-Frequency:	15%, 36
Rare:	4%, 10
Suitability for general audience score:	89
Number Of Words:	246

Student's work (excerpt): minimal adaption needed for non-experts

Rhythm disorder of the heart is a proximal cause for heart failure. Nowadays treatment for rhythmic disorder relay on electronic pacemakers. Although an excellent solution, it withholds disadvantages such as the need for battery change, and the risk of contaminations. Thus, a biological alternative may be ideal. Stem cells that are generated from the person's own hair, and can differentiate in to heart-like cells, hold a therapeutic promise.

To assess compatibility of these heart-like cells to human physiology, their functionality should be investigated. Hence, our major goal is to characterize their electrical behavior and examine whether they functionally recapitulate adult human heart cells.

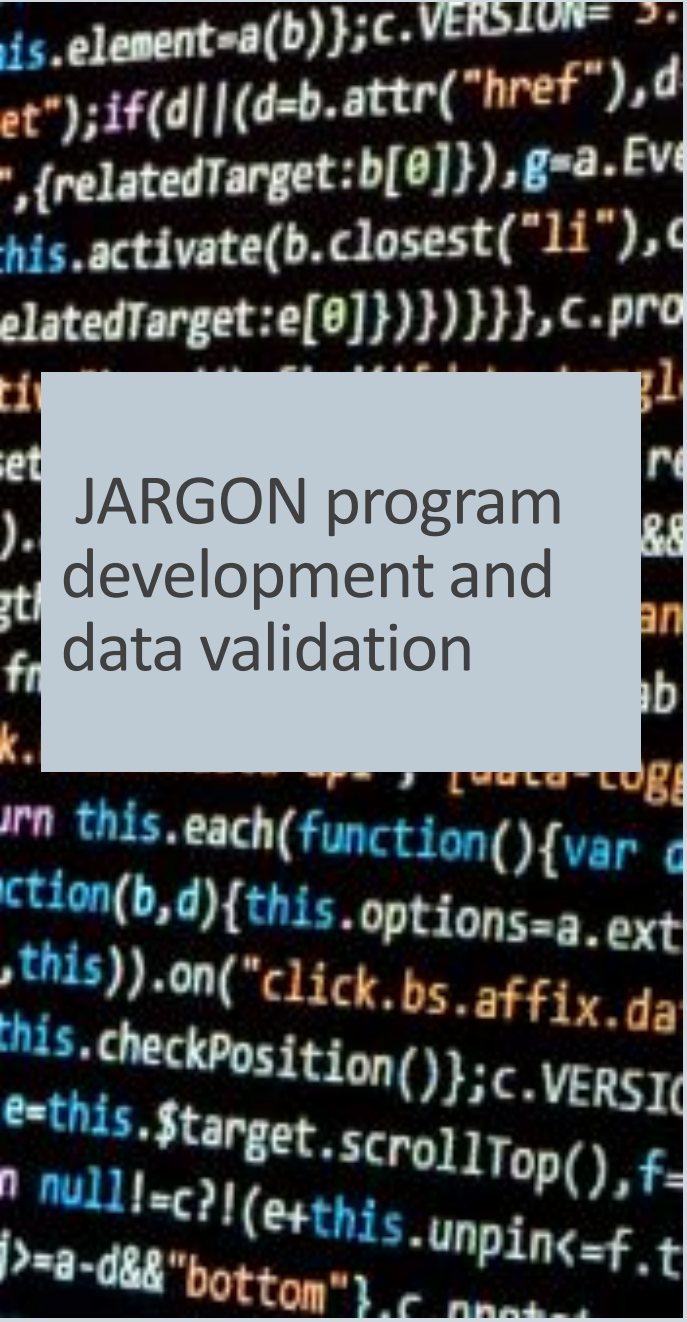
Although much research is still due, this novel biological solution allows optimistic hopes.

Common:	84%, 99
Mid-Frequency:	12%, 14
Rare:	4%, 5
Suitability for general audience score:	90
Number Of Words:	118

Student's work (excerpt): well-adapted for non-experts

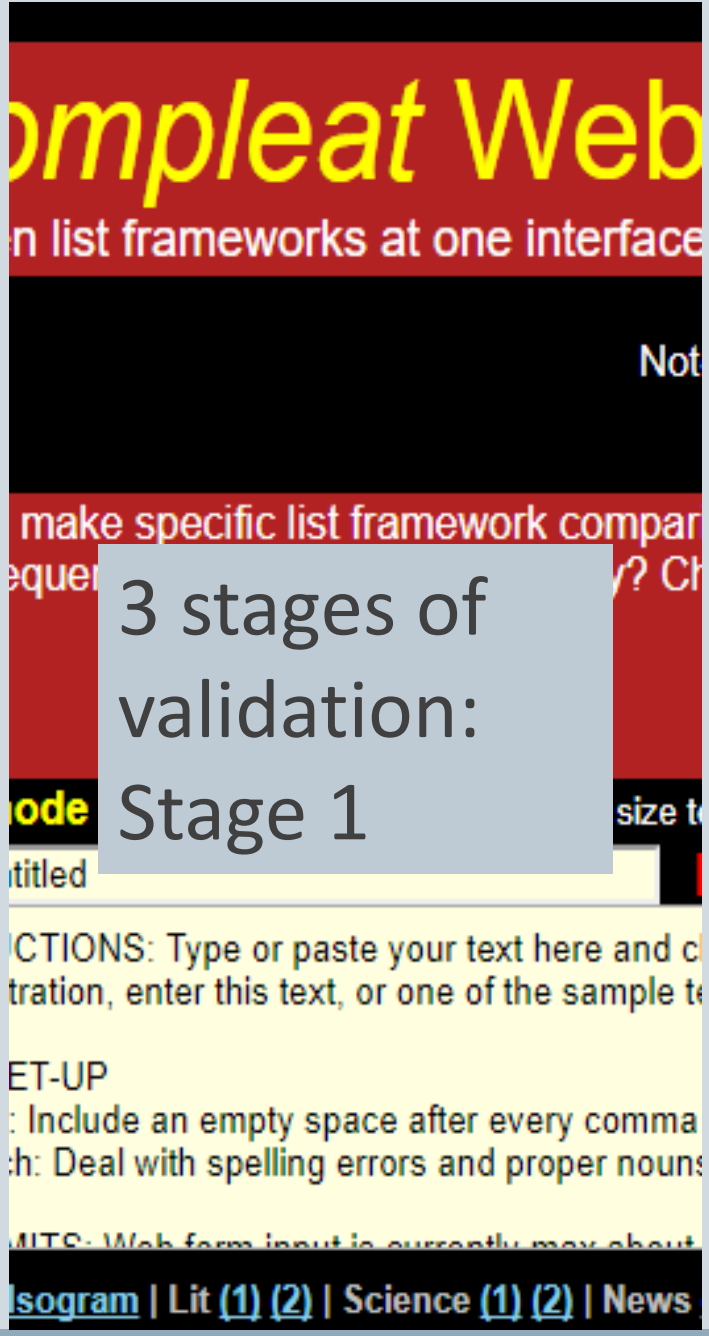
Interrupted trillion times a day? That's may make you more creative!
Trying not to **sink** in a flood of e-**mails**, text messages and phone calls, I tried to work on my master **thesis** in **behavioral sciences** and management at the **Technion**. If the following research I managed to perform under this **interruptions** attack sounds surprising to you, it **proves** its own argument: **interruptions** can make us more creative.
Numerous studies in **psychology** have been busy in recent years trying to answer whether **interruptions** are bad to our performance and mood. Actually, most of these studies answered a different question: how much are **interruptions** bad.
But these studies **overlooked** a certain type of tasks, in which **forgetting** of what you have done a second ago is helpful: creative tasks! When trying to come up with a new idea, one is often stuck on too **predictable** ways of thought. An **interruption** may help one to turn over to a new **leaf**.
We tested this **hypothesis** in a lab study. We gave 61 students two creative tasks. One third of the participants worked on the tasks **continuously**, without any **interruption**. The rest were **interrupted** during the tasks by simple **arithmetic exercises** that **popped**-up on their screen, either once for a longer period, or multiple times. We found that those participants who were **interrupted** were more creative, and specifically their creative performance improved after each **interruption**. So next time you are thinking about turning off your phone in order to concentrate on a task, think again. If it's a creative task, keep it on and let yourself be inspired by **interruptions**.

Common:	89%, 241
Mid-Frequency:	10%, 28
Rare:	0%, 1
Suitability for general audience score:	94



5 pilot stages of development: final stage:

- ~250,000 articles published in the BBC sites during the years 2012-2015
- Over 90 million words counted using a crawler
- ~500,000 word types were ordered by number of appearances



Results from BNC-COCA VPcompleat program was compared:

- Quantitatively and qualitatively on students' writing pre/posttests from an academic writing course (Rakedzon & Baram-Tsabari, 2017)

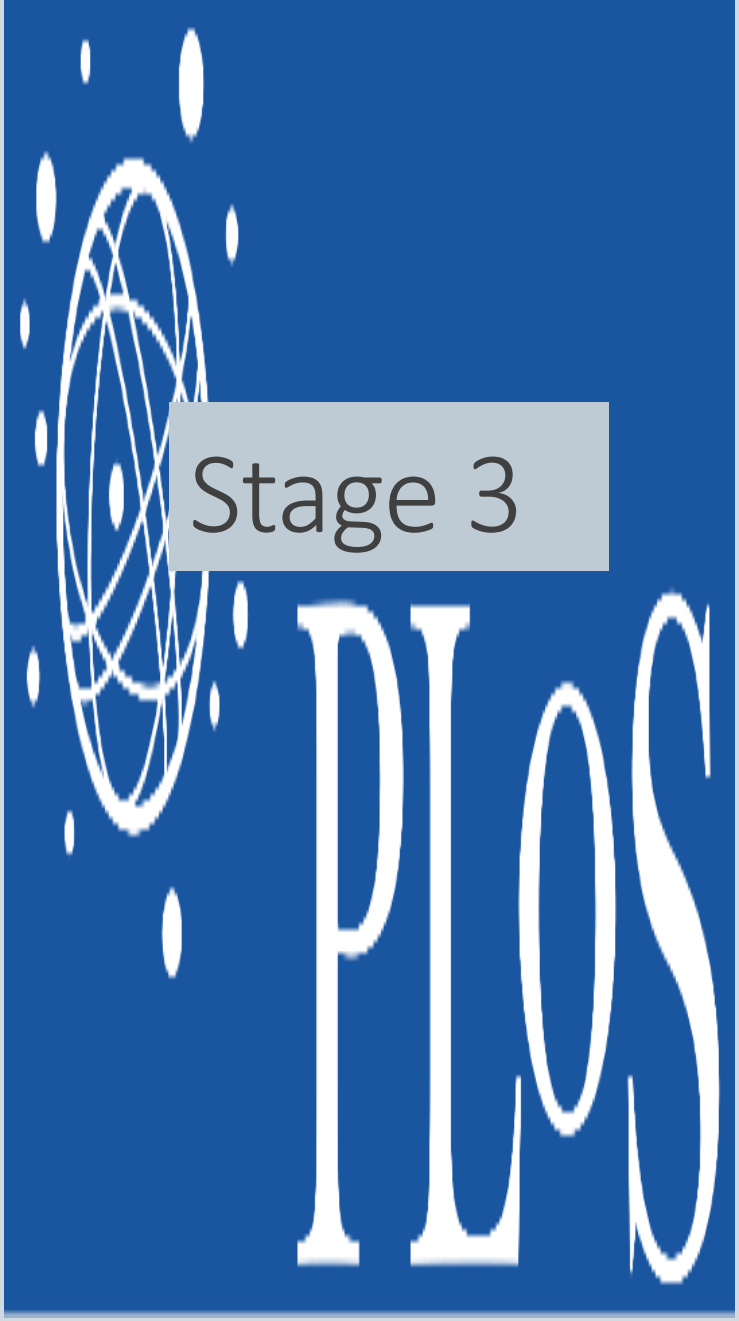
- Validity of the comparison -> Pearson Correlation

https://www.lextutor.ca/vp/comp/ (Nation, 2012; Cobb, 2016)

Comparing jargon use:

Results from the De-jargonizer vs. results from Sharon & Baram-Tsabari (2013) using transcripts of non-science lectures, TED science lectures and academic scientific lectures

Stage 3



5000 pairs: published research abstracts and lay reader summaries describing the same article were compared

- From PLoS computational biology and PLoS genetics journals

How can the De-jargonizer be used:

1. in the classroom?
2. in research?

Automatic jargon identifier for scientists engaging with the public and science communication educators (Rakedzon et al, 2017)

Vocabulary has been implicated in reading comprehension in general

Aim

To use the jargon identifier tool to test and evaluate students' use of jargon in pre and posttests from an academic writing course for graduate STEM students with an intervention lesson on science communication.

Research questions

1. How much jargon do graduate STEM students use when describing their own research to lay audience and to the academic community?

2. How does a one lesson intervention about science communication embedded in a compulsory academic writing course affect students' use of jargon?

Methods: Setting

The academic writing course: a 14-week compulsory course for PhD students

- Some prior knowledge of IMRAD and strong basic English skills
- Conducted in English by native speakers
- Weekly lectures/exercises; student-teacher conferences; 1 pre/post task and 4 formative assessment tasks with feedback

Participants /sample

Technion PhD students:

- 888 writing samples from 222 students' pre/posttests to test jargon

Diverse student body:

- Hebrew (~70%), Russian (13%), Arabic (5%)
- Evenly spread by gender
- All STEM faculties

Common:

Instruments

Rare:

Suitability for general audience

Pre/posttest for assessing language and genre:

Describe your research project (150-250 words each)

1. For the academic community/academic journal
2. For the general public

- Analyzed by the De-jargonizer
- Paired samples T-tests for comparing texts

Procedure

	Pre test	Academic writing course	Popular lecture intervention	Popular task intervention	Post test
Control (no course)	X				X
Comparison (course w/o intervention)	X	X			X
Popular science lecture **intervention**	X	X	X		X
Popular science lecture + task **intervention**	X	X	X	X	X

Results

Analysis of both intervention groups showed similar results

Academic writing (*p<0.05)

Word type	Pre-post change in use
High frequency	Decrease* ↘
Mid-frequency	Increase* ↗
Jargon	Increase* ↗

Results

Popular science writing

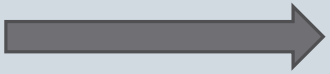
(*p<0.05)

Word type	Pre-post change in use
High frequency	Decrease* ↘
Mid-frequency	Increase* ↗
Jargon	No change

Results

- Popular science writing pre-posttests: **~6%** jargon
- Academic writing pre/posttests **~10%** jargon

- Significantly lower percentage of jargon between genres



indicating some awareness of adaptation for lay audiences

Discussion

Graduate students are aware of the appropriate jargon in their academic writing: average of ~10% of the words are jargon

Discussion

For popular science writing, jargon was a confusing issue

- Use less jargon in the popular as opposed academic task
- Not always enough for the general audience
- Our research showed students more easily use explanation instead of synonyms (Rakedzon & Baram-Tsabari, 2017)

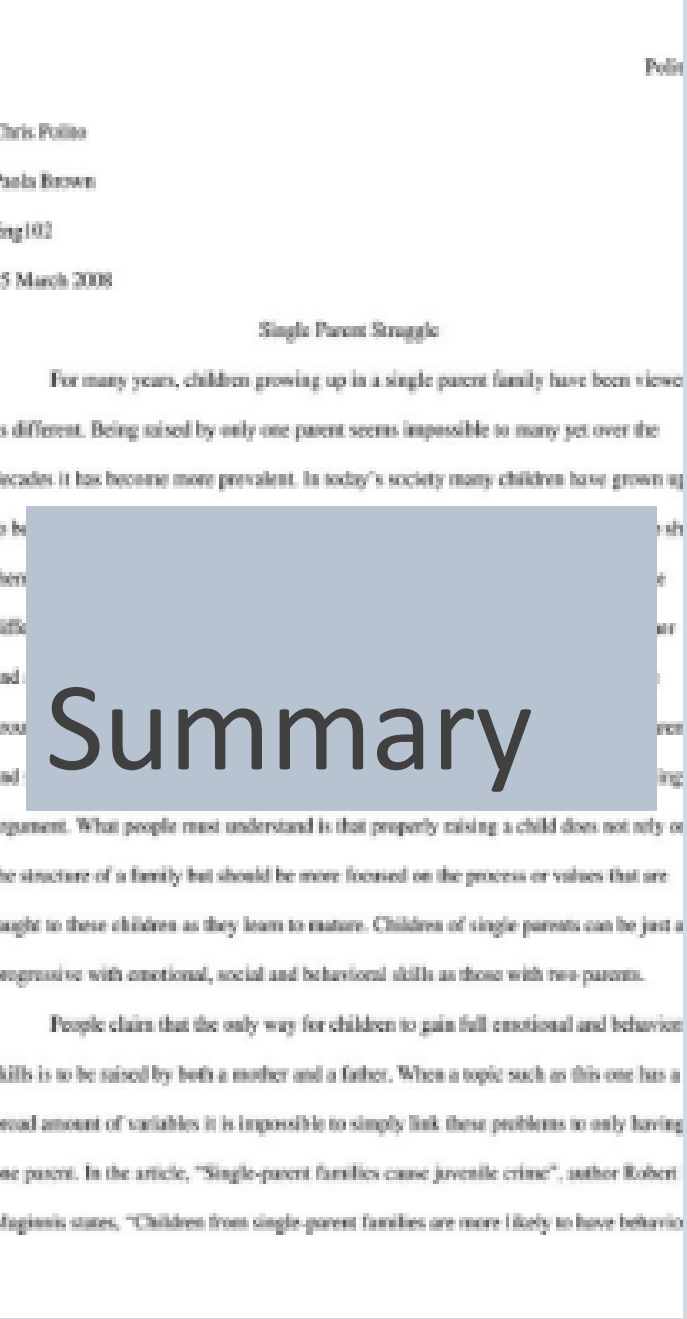
Discussion

Reason for too much jargon:

- May be from extremely complicated nature of much of the research at the Technion
- Need for longer intervention/ more instruction

Further research could look at:

- Offer solutions to problematic words
- Longer interventions
- Comparing
  - L1 (native) to L2 writers
  - Different disciplines
  - Students of different degrees
  - Scientists/professionals at different stages of their career
  - Different genres



The De-jargonizer

- Measures jargon in a color-coded, user-friendly way
- Good agreement with results from other studies (Sharon & Baram-Tsabari, 2013; Nation, 2012)
- Can be used by educators in assessment of texts, vocabulary and communication training
- Can also aid researchers in testing written and spoken texts

De-Jargonizer

Thank you for listening!

Please visit: **http://scienceandpublic**.com/

More about the project:

Rakedzon, T. & **Baram-Tsabari**, A. (2017) **Assessing** and improving L2 graduate students' popular science and academic writing in an academic writing course. Educational **Psychology**.

Rakedzon, T. & **Baram-Tsabari**, A. (2017) To make a long story short: A **rubric** for **assessing** graduate written science communication. **Assessing** writing.

Rakedzon, T., **Segev**, E., **Baram-Tsabari**, A., **Chapnik**, N., **Yosef**, R. (2017) Automatic **jargon identifier** for scientists **engaging** with the public and science communication **educators**. **PLoS** ONE.

hutzipi@technion.ac.il

Your texts.....

- Open the site http://scienceandpublic.com/

- You can also choose the Hebrew site in the upper bar
- Upload a WORD file

**At the moment, you can easily save your results with a screenshot or snip shot