Developing an in-corpus and high-frequency word list(s) for science majors

Suwako Uehara

8 July 2022

PanSIG
University of Nagano
July 8-10, 2022

Literature Review

Lack of vocabulary leads to lower comprehension (Kelly, 1991)

95% lexical coverage - learners need help with vocab for comprehension

–98% lexical coverage - learners reading comprehension should be okay (Nation, 2006; Webb & Macalister, 2012; Webb & Rodgers, 2009; He & Godfroid, 2019)

Literature Review

Corpus (Biber et al, 1998)

High-frequency word list (Xue & Nation, 1984)

Discipline-specific corpus (e.g. Coxhead & Hirsch, 2000; Ward, 2009)

Lit. Review: Science Related Corpus and Word Lists

Discipline	Summary of University Focused Corpus	Corpus Size	Frequency Word List	Reference
Engineering	25 textbook recommendations commonly used for 3rd-4th year undergraduate students	271,000 words	299-word list for foundation engineering by flemma	Ward (2009)
Science, engineering, technology (22 domain)	Corpus of academic papers across 22 domains in science and engineering with high impact factors (http://www.perc21.org/cpe_project/index.html)	17 million words	1260 word families Headwords	Nesi (2012)
Science	Reading materials (textbooks, lecture notes) for 1st year students across 14 science subjects (e.g. Agricultural science, Biology Chemistry, physics, Mathematics, Computer Science etc.)	1.76 million words	315 word families	Coxhead & Hirsch (2007)
Engineering	Compulsory engineering textbooks	2 million words	8850 word-types	Mundraya (2006)
Science and Engineering	JIaotong Daxue English of science and technology (JDEST) corpus	1 million words		Yang (1986)
Engineering	Student Engineering English Corpus (SEEC)	2 million words		Moudraia (2003; 2004)
Science, Engineering, Social Sciences	Academic corpus from 30 research articles, seven textbook chapters, 20 academic book reviews in each of seven disciplines; 45 scientific letters in physics and biology theses, research articles, 8 Master's thesis, six doctoral dissertations, 8 final year BSc thesis across six disciplines	3 million words		Hyland and Tse (2007)

Work in Progress

Aim: Derive a high-frequency word list from a corpus of professor recommended reading materials for Graduate school science and engineering students in a department of Engineering Science at a national university.

Programs: Electronic Engineering, Optical Science and Engineering, Applied Physics, and Chemistry and Biotechnology

Decisions required to make a corpus and high frequency word list(s)

-Representation of the Corpus (Sinclair, 1991): Science Engineering prof recommended documents

-Corpus size (Coxhead & Hirsch, 2007): 1 million words

-High-frequency word list size (1000 words)

-Word types: Flemma (McLean, 2018)

-Removal of NGSL (Coxhead & Hirsch, 2007): 1st-2nd 1000 NGSL

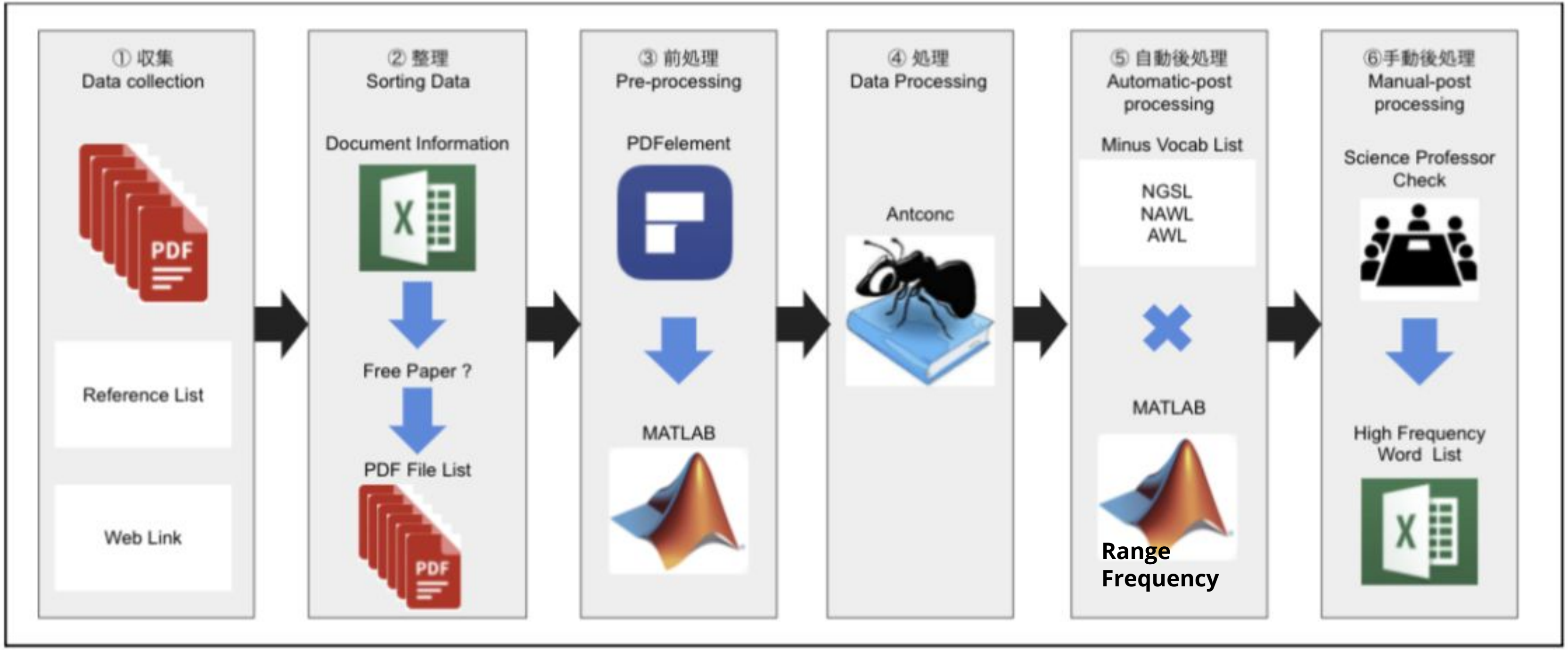
Research Question

RQ1 What kind of vocabulary list(s) would benefit science major students in graduate school?

RQ2 What method will make corpus and word frequency list for science and engineering students efficient?

RQ3 What kind of decisions are required to creating a “clean” corpus?

Method: Process to create a discipline specific corpus and high-frequency word list



Method (Work in Progress)

Data set

Received recommendations (*N* = 22 profs; *N* = 1181 docs) in the form of pdf, reference list, & weblinks.

Data processed for today’s presentation

(*N* = 10 profs; *N* = 330 docs)

① 収集
Data collection
PDF
Reference List
Web Link

Method: Data Set Recommended from Science and Engineering Department

Program	Profs	Docs	Types of data (J. Impact factor)
Chemistry and Biotechnology	10	300	Am J Physiol Regul Intergr Comp Phys (3.62); Amino Acids (3.23); Anal. Chem (3.23); Annu. Rev. Biochemical (23.64); Artificial DNA: PNA & XNA (-); Biochemical and Biophysical Research Communications (3.58); Bioconjugate Chem. (38.77); Biomed. Opt. Express (3.73); Biomol. Chem (3.88); Biopolymers (Peptide Science) (2.51); Cell (41.58); Chem. Eur. J. (5.24); Chem. Ur. J (5.23); ChemBioChem (3.16); Current Biology (10.83); Current Pharmaceutical Design (2.21); FEBS open bio (2.69); FEMS Microbiology Reviews (16.41); J Physiol (5.18); J. Am. Soc. Mass Spectrum (3.11); J. Phys. Chem B (2.99); J.Chem. Phys. (3.49); Mol. BioSyst. 3.34); Moleciular Microbiology (3.82); Molecules (4.15); Nanomaterials (4.03); Nat rev Microbiol (60.63); Nature (49.96); Nature Chem (24.23); Nature Communications (14.92); Org. Biomol. Chem (3.88); Org. Biomol. Chem (3.88); Science (33.61); Scientific Reports (4.525); Soft matter (3.68); Soft matter (3.68); The Chemical Society of Japan (5.49); University Doctorate dissertations (-)

Method: Data Set recommended from Chemistry and BioTechnology

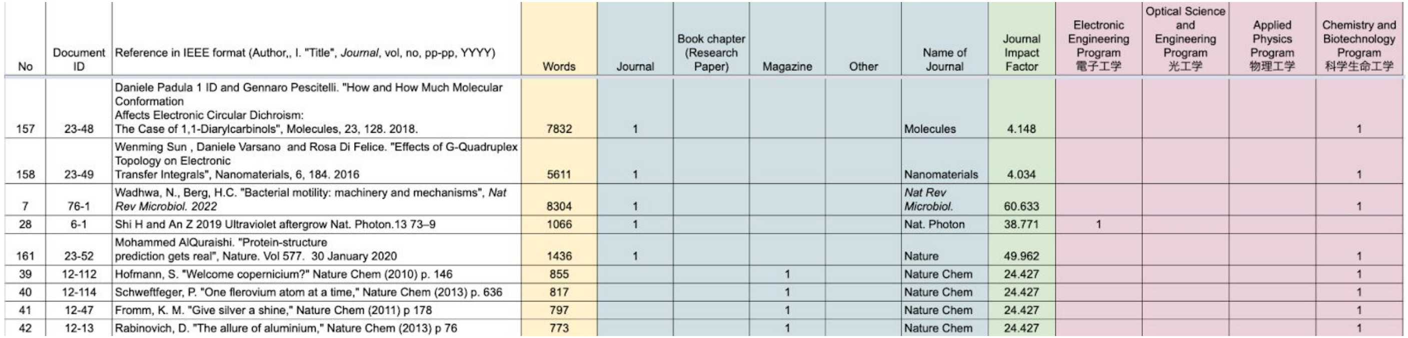
Type of Reading Material	*n*	Profs	Additional information about the reading materials
Journals	238	10	Research articles published by the professors that recommended them, or papers that are relevant to the professor's lab, or papers that are highly cited in the field. Impact factor range 3.23–49.96 ***Uneven distribution of papers from each professor (low: 4, high: 188)** **Spoiler alert: High frequency word list distorted from the high 188 recommendation from one lab.**
Book Chapter	1	1	Book chapter recommend from one research lab
Magazine articles	89	1	Short articles with reading materials from N*ature Chemistry*,with an impact factor of 24.427
Doctorate Dissertation	2	1	Dissertations highly connected to research in the professors' research lab.
Total			

② Data Sorting

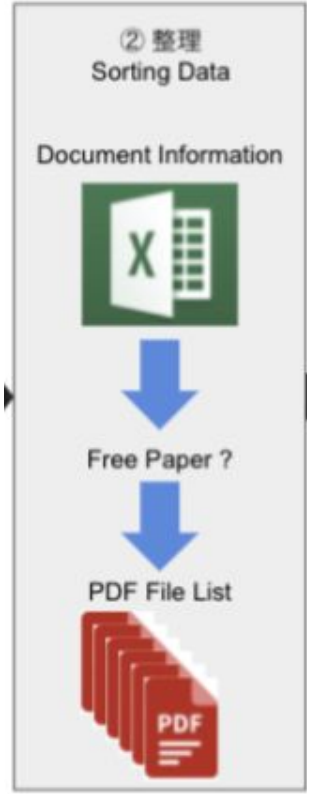
- Created a list of documents

Teacher ID and doc no.; Reference; Type of document; Word count; Journal Impact Factor

- Saved PDF documents



No	Document ID	Reference in IEEE format (Author,, I. "Title", *Journal*, vol, no, pp-pp, YYYY)	Words	Journal	Book chapter (Research Paper)	Magazine	Other	Name of Journal	Journal Impact Factor	Electronic Engineering Program 電子工学	Optical Science and Engineering Program 光工学	Applied Physics Program 物理工学	Chemistry and Biotechnology Program 科学生命工学
157	23-48	Daniele Padula 1 ID and Gennaro Pescitelli. "How and How Much Molecular Conformation Affects Electronic Circular Dichroism: The Case of 1,1-Diarylcarbinols", Molecules, 23, 128. 2018.	7832	1				Molecules	4.148				1
158	23-49	Wenming Sun , Daniele Varsano and Rosa Di Felice. "Effects of G-Quadruplex Topology on Electronic Transfer Integrals", Nanomaterials, 6, 184. 2016	5611	1				Nanomaterials	4.034				1
7	76-1	Wadhwa, N., Berg, H.C. "Bacterial motility: machinery and mechanisms", *Nat Rev Microbiol. 2022*	8304	1				*Nat Rev Microbiol.*	60.633				1
28	6-1	Shi H and An Z 2019 Ultraviolet aftergrow Nat. Photon.13 73–9	1066	1				Nat. Photon	38.771	1			
161	23-52	Mohammed AlQuraishi. "Protein-structure prediction gets real", Nature. Vol 577. 30 January 2020	1436	1				Nature	49.962				1
39	12-112	Hofmann, S. "Welcome copernicium?" Nature Chem (2010) p. 146	855			1		Nature Chem	24.427				1
40	12-114	Schweftfeger, P. "One flerovium atom at a time," Nature Chem (2013) p. 636	817			1		Nature Chem	24.427				1
41	12-47	Fromm, K. M. "Give silver a shine," Nature Chem (2011) p 178	797			1		Nature Chem	24.427				1
42	12-13	Rabinovich, D. "The allure of aluminium," Nature Chem (2013) p 76	773			1		Nature Chem	24.427				1



② 整理
Sorting Data

Document Information

Free Paper ?

PDF File List

③ Pre-Processing

Edited text to keep only data required for processing

Kept

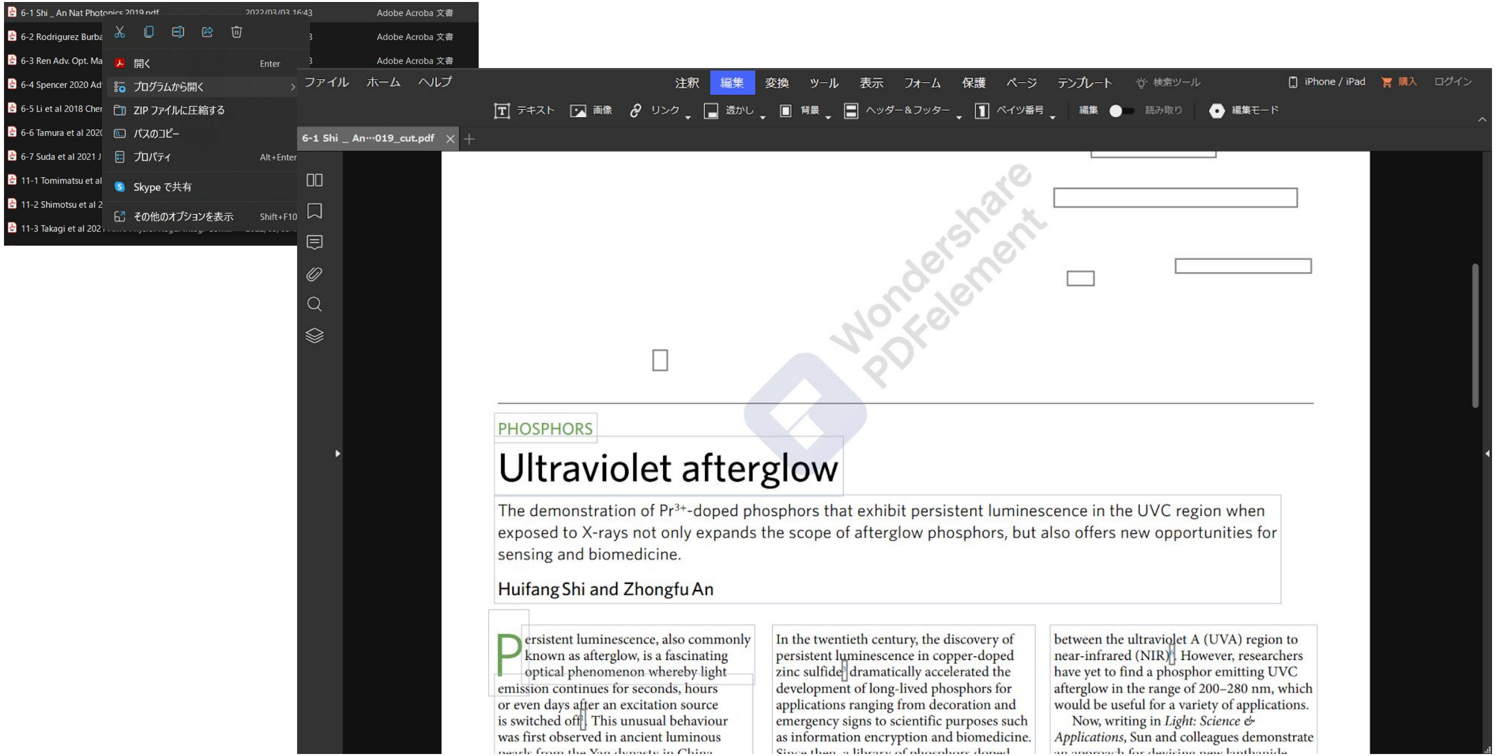
- Title
- Abstract
- Introduction
- Method
- Experimental Section
- Results
- Discussion
- Conclusion
- Figure and Table caption

Removed

- Author name
- Figures
- Tables
- Acknowledgements
- Reference list
- Headers
- Footers
- Page Numbers
- Links
- Journal Names
- Journal Library
- E-mail addresses
- Stand-alone formula (E.g. equations that is not in a text)
- Images (e.g. Journal logo, search engine logo, e.t.c.)

③ 前処理
Pre-processing
PDFelement
MATLAB

③ Pre-Processing: PDFelement



6-1 Shi _ An Nat Photonics 2019.pdf 2022/03/03 16:43 Adobe Acroba 文書
6-2 Rodrigurez Burba … Adobe Acroba 文書
6-3 Ren Adv. Opt. Ma … Adobe Acroba 文書
6-4 Spencer 2020 Ad
6-5 Li et al 2018 Cher
6-6 Tamura et al 2020
6-7 Suda et al 2021 J
11-1 Tomimatsu et al
11-2 Shimotsu et al 2
11-3 Takagi et al 2021

PHOSPHORS

Ultraviolet afterglow

The demonstration of Pr^{3+}-doped phosphors that exhibit persistent luminescence in the UVC region when exposed to X-rays not only expands the scope of afterglow phosphors, but also offers new opportunities for sensing and biomedicine.

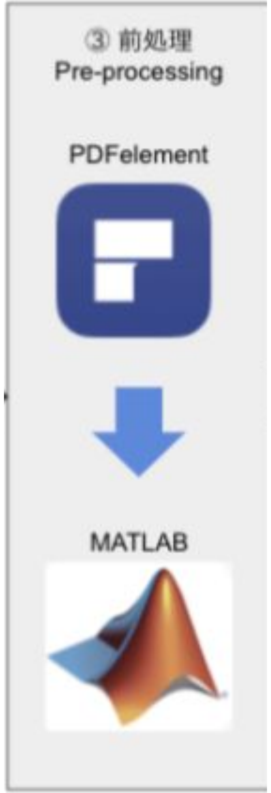
Huifang Shi and Zhongfu An

Persistent luminescence, also commonly known as afterglow, is a fascinating optical phenomenon whereby light emission continues for seconds, hours or even days after an excitation source is switched off. This unusual behaviour was first observed in ancient luminous pearls from the Yan dynasty in China

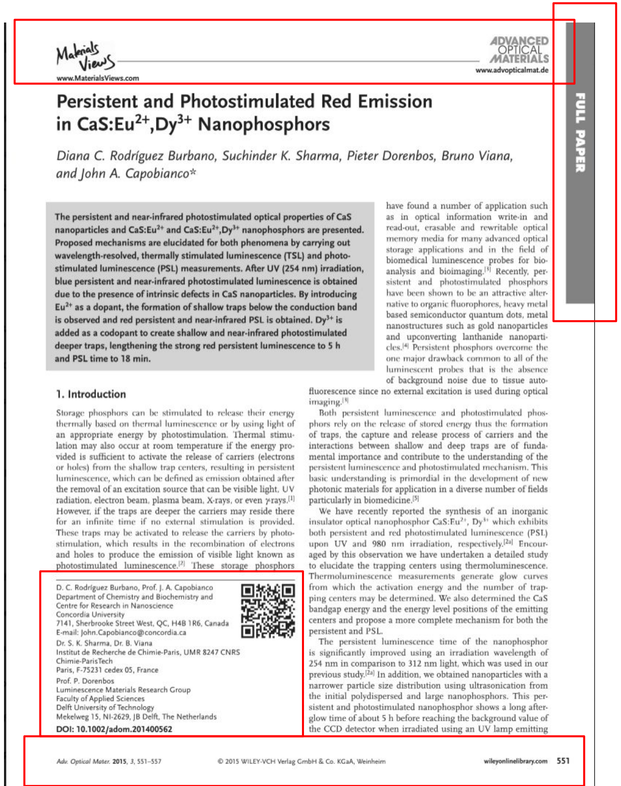
In the twentieth century, the discovery of persistent luminescence in copper-doped zinc sulfide dramatically accelerated the development of long-lived phosphors for applications ranging from decoration and emergency signs to scientific purposes such as information encryption and biomedicine. Since then, a library of phosphors doped

between the ultraviolet A (UVA) region to near-infrared (NIR). However, researchers have yet to find a phosphor emitting UVC afterglow in the range of 200–280 nm, which would be useful for a variety of applications.

Now, writing in *Light: Science & Applications*, Sun and colleagues demonstrate an approach for devising new lanthanide



Step 3 Pre-Processing



Materials Views
www.MaterialsViews.com

ADVANCED OPTICAL MATERIALS
www.advopticalmat.de

FULL PAPER

Persistent and Photostimulated Red Emission in CaS:Eu^{2+},Dy^{3+} Nanophosphors

*Diana C. Rodríguez Burbano, Suchinder K. Sharma, Pieter Dorenbos, Bruno Viana, and John A. Capobianco**

The persistent and near-infrared photostimulated optical properties of CaS nanoparticles and CaS:Eu^{2+} and CaS:Eu^{2+},Dy^{3+} nanophosphors are presented. Proposed mechanisms are elucidated for both phenomena by carrying out wavelength-resolved, thermally stimulated luminescence (TSL) and photostimulated luminescence (PSL) measurements. After UV (254 nm) irradiation, blue persistent and near-infrared photostimulated luminescence is obtained due to the presence of intrinsic defects in CaS nanoparticles. By introducing Eu^{2+} as a dopant, the formation of shallow traps below the conduction band is observed and red persistent and near-infrared PSL is obtained. Dy^{3+} is added as a codopant to create shallow and near-infrared photostimulated deeper traps, lengthening the strong red persistent luminescence to 5 h and PSL time to 18 min.

1. Introduction

Storage phosphors can be stimulated to release their energy thermally based on thermal luminescence or by using light of an appropriate energy by photostimulation. Thermal stimulation may also occur at room temperature if the energy provided is sufficient to activate the release of carriers (electrons or holes) from the shallow trap centers, resulting in persistent luminescence, which can be defined as emission obtained after the removal of an excitation source that can be visible light, UV radiation, electron beam, plasma beam, X-rays, or even γ-rays.[1] However, if the traps are deeper the carriers may reside there for an infinite time if no external stimulation is provided. These traps may be activated to release the carriers by photostimulation, which results in the recombination of electrons and holes to produce the emission of visible light known as photostimulated luminescence.[2] These storage phosphors have found a number of application such as in optical information write-in and read-out, erasable and rewritable optical memory media for many advanced optical storage applications and in the field of biomedical luminescence probes for bioanalysis and bioimaging.[3] Recently, persistent and photostimulated phosphors have been shown to be an attractive alternative to organic fluorophores, heavy metal based semiconductor quantum dots, metal nanostructures such as gold nanoparticles and upconverting lanthanide nanoparticles.[4] Persistent phosphors overcome the one major drawback common to all of the luminescent probes that is the absence of background noise due to tissue autofluorescence since no external excitation is used during optical imaging.[3]

Both persistent luminescence and photostimulated phosphors rely on the release of stored energy thus the formation of traps, the capture and release process of carriers and the interactions between shallow and deep traps are of fundamental importance and contribute to the understanding of the persistent luminescence and photostimulated mechanism. This basic understanding is primordial in the development of new photonic materials for application in a diverse number of fields particularly in biomedicine.[5]

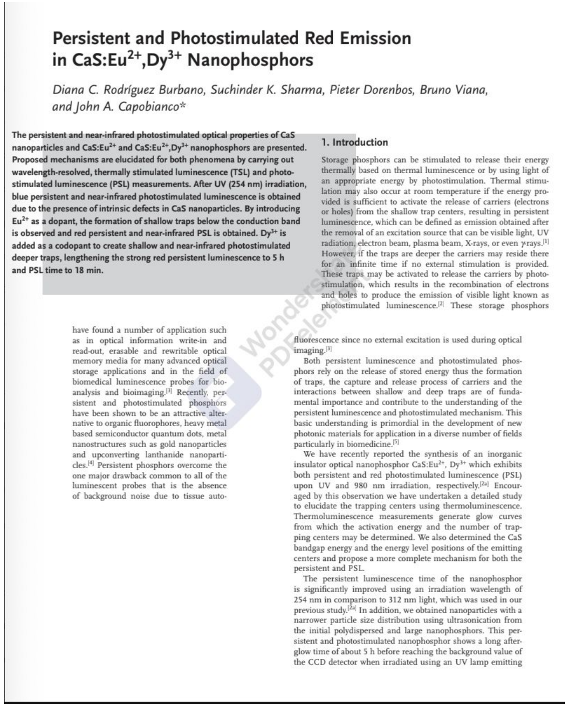
We have recently reported the synthesis of an inorganic insulator optical nanophosphor CaS:Eu^{2+}, Dy^{3+} which exhibits both persistent and red photostimulated luminescence (PSL) upon UV and 980 nm irradiation, respectively.[2a] Encouraged by this observation we have undertaken a detailed study to elucidate the trapping centers using thermoluminescence. Thermoluminescence measurements generate glow curves from which the activation energy and the number of trapping centers may be determined. We also determined the CaS bandgap energy and the energy level positions of the emitting centers and propose a more complete mechanism for both the persistent and PSL.

The persistent luminescence time of the nanophosphor is significantly improved using an irradiation wavelength of 254 nm in comparison to 312 nm light, which was used in our previous study.[2a] In addition, we obtained nanoparticles with a narrower particle size distribution using ultrasonication from the initial polydispersed and large nanophosphors. This persistent and photostimulated nanophosphor shows a long afterglow time of about 5 h before reaching the background value of the CCD detector when irradiated using an UV lamp emitting

D. C. Rodríguez Burbano, Prof. J. A. Capobianco
Department of Chemistry and Biochemistry and
Centre for Research in Nanoscience
Concordia University
7141, Sherbrooke Street West, QC, H4B 1R6, Canada
E-mail: John.Capobianco@concordia.ca
Dr. S. K. Sharma, Dr. B. Viana
Institut de Recherche de Chimie-Paris, UMR 8247 CNRS
Chimie-ParisTech
Paris, F-75231 cedex 05, France
Prof. P. Dorenbos
Luminescence Materials Research Group
Faculty of Applied Sciences
Delft University of Technology
Mekelweg 15, NI-2629, JB Delft, The Netherlands

DOI: 10.1002/adom.201400562

Before pre-processing



Persistent and Photostimulated Red Emission in CaS:Eu^{2+},Dy^{3+} Nanophosphors

Diana C. Rodríguez Burbano, Suchinder K. Sharma, Pieter Dorenbos, Bruno Viana, and John A. Capobianco*

The persistent and near-infrared photostimulated optical properties of CaS nanoparticles and CaS:Eu^{2+} and CaS:Eu^{2+},Dy^{3+} nanophosphors are presented. Proposed mechanisms are elucidated for both phenomena by carrying out wavelength-resolved, thermally stimulated luminescence (TSL) and photostimulated luminescence (PSL) measurements. After UV (254 nm) irradiation, blue persistent and near-infrared photostimulated luminescence is obtained due to the presence of intrinsic defects in CaS nanoparticles. By introducing Eu^{2+} as a dopant, the formation of shallow traps below the conduction band is observed and red persistent and near-infrared PSL is obtained. Dy^{3+} is added as a codopant to create shallow and near-infrared photostimulated deeper traps, lengthening the strong red persistent luminescence to 5 h and PSL time to 18 min.

1. Introduction

Storage phosphors can be stimulated to release their energy thermally based on thermal luminescence or by using light of an appropriate energy by photostimulation. Thermal stimulation may also occur at room temperature if the energy provided is sufficient to activate the release of carriers (electrons or holes) from the shallow trap centers, resulting in persistent luminescence, which can be defined as emission obtained after the removal of an excitation source that can be visible light, UV radiation, electron beam, plasma beam, X-rays, or even γ-rays.[1] However, if the traps are deeper the carriers may reside there for an infinite time if no external stimulation is provided. These traps may be activated to release the carriers by photostimulation, which results in the recombination of electrons and holes to produce the emission of visible light known as photostimulated luminescence.[2] These storage phosphors have found a number of application such as in optical information write-in and read-out, erasable and rewritable optical memory media for many advanced optical storage applications and in the field of biomedical luminescence probes for bioanalysis and bioimaging.[3] Recently, persistent and photostimulated phosphors have been shown to be an attractive alternative to organic fluorophores, heavy metal based semiconductor quantum dots, metal nanostructures such as gold nanoparticles and upconverting lanthanide nanoparticles.[4] Persistent phosphors overcome the one major drawback common to all of the luminescent probes that is the absence of background noise due to tissue autofluorescence since no external excitation is used during optical imaging.[3]

Both persistent luminescence and photostimulated phosphors rely on the release of stored energy thus the formation of traps, the capture and release process of carriers and the interactions between shallow and deep traps are of fundamental importance and contribute to the understanding of the persistent luminescence and photostimulated mechanism. This basic understanding is primordial in the development of new photonic materials for application in a diverse number of fields particularly in biomedicine.[5]

We have recently reported the synthesis of an inorganic insulator optical nanophosphor CaS:Eu^{2+}, Dy^{3+} which exhibits both persistent and red photostimulated luminescence (PSL) upon UV and 980 nm irradiation, respectively.[2a] Encouraged by this observation we have undertaken a detailed study to elucidate the trapping centers using thermoluminescence. Thermoluminescence measurements generate glow curves from which the activation energy and the number of trapping centers may be determined. We also determined the CaS bandgap energy and the energy level positions of the emitting centers and propose a more complete mechanism for both the persistent and PSL.

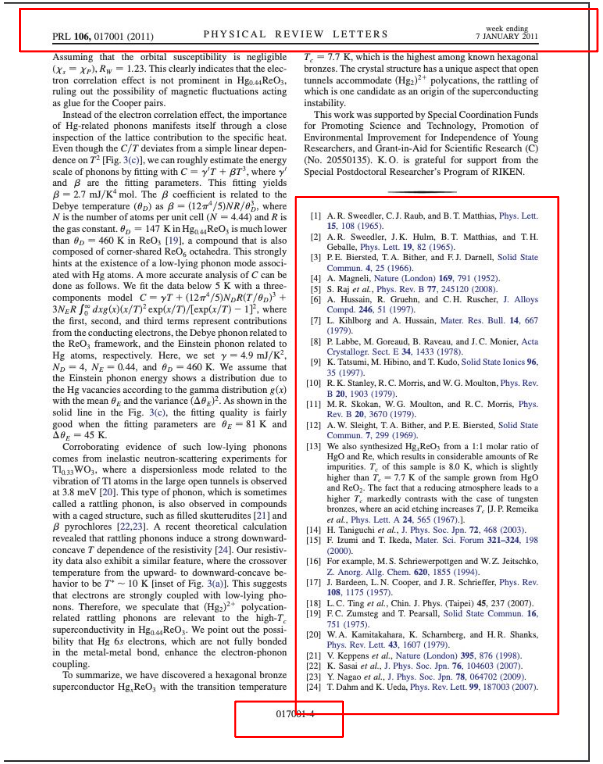
The persistent luminescence time of the nanophosphor is significantly improved using an irradiation wavelength of 254 nm in comparison to 312 nm light, which was used in our previous study.[2a] In addition, we obtained nanoparticles with a narrower particle size distribution using ultrasonication from the initial polydispersed and large nanophosphors. This persistent and photostimulated nanophosphor shows a long afterglow time of about 5 h before reaching the background value of the CCD detector when irradiated using an UV lamp emitting

After pre-processing

③ 前処理
Pre-processing
PDFelement
MATLAB

(6-2 Rodriguez)

Step 3 Pre-Processing



PRL **106,** 017001 (2011) PHYSICAL REVIEW LETTERS week ending 7 JANUARY 2011

Assuming that the orbital susceptibility is negligible ($\chi_s = \chi_P$), $R_W = 1.23$. This clearly indicates that the electron correlation effect is not prominent in $Hg_{0.44}ReO_3$, ruling out the possibility of magnetic fluctuations acting as glue for the Cooper pairs.

Instead of the electron correlation effect, the importance of Hg-related phonons manifests itself through a close inspection of the lattice contribution to the specific heat. Even though the C/T deviates from a simple linear dependence on T^2 [Fig. 3(c)], we can roughly estimate the energy scale of phonons by fitting with $C = \gamma' T + \beta T^3$, where γ' and β are the fitting parameters. This fitting yields $\beta = 2.7$ mJ/K^4 mol. The β coefficient is related to the Debye temperature (θ_D) as $\beta = (12\pi^4/5)NR/\theta_D^3$, where N is the number of atoms per unit cell ($N = 4.44$) and R is the gas constant. $\theta_D = 147$ K in $Hg_{0.44}ReO_3$ is much lower than $\theta_D = 460$ K in ReO_3 [19], a compound that is also composed of corner-shared ReO_6 octahedra. This strongly hints at the existence of a low-lying phonon mode associated with Hg atoms. A more accurate analysis of C can be done as follows. We fit the data below 5 K with a three-components model $C = \gamma T + (12\pi^4/5)N_D R(T/\theta_D)^3 + 3N_E R \int_0^\infty dx g(x)(x/T)^2 \exp(x/T)/[\exp(x/T) - 1]^2$, where the first, second, and third terms represent contributions from the conducting electrons, the Debye phonon related to the ReO_3 framework, and the Einstein phonon related to Hg atoms, respectively. Here, we set $\gamma = 4.9$ mJ/K^2, $N_D = 4$, $N_E = 0.44$, and $\theta_D = 460$ K. We assume that the Einstein phonon energy shows a distribution due to the Hg vacancies according to the gamma distribution $g(x)$ with the mean θ_E and the variance $(\Delta\theta_E)^2$. As shown in the solid line in the Fig. 3(c), the fitting quality is fairly good when the fitting parameters are $\theta_E = 81$ K and $\Delta\theta_E = 45$ K.

Corroborating evidence of such low-lying phonons comes from inelastic neutron-scattering experiments for $Tl_{0.33}WO_3$, where a dispersionless mode related to the vibration of Tl atoms in the large open tunnels is observed at 3.8 meV [20]. This type of phonon, which is sometimes called a rattling phonon, is also observed in compounds with a caged structure, such as filled skutterudites [21] and β pyrochlores [22,23]. A recent theoretical calculation revealed that rattling phonons induce a strong downward-concave T dependence of the resistivity [24]. Our resistivity data also exhibit a similar feature, where the crossover temperature from the upward- to downward-concave behavior to be $T^* \sim 10$ K [inset of Fig. 3(a)]. This suggests that electrons are strongly coupled with low-lying phonons. Therefore, we speculate that $(Hg_2)^{2+}$ polycation-related rattling phonons are relevant to the high-T_c superconductivity in $Hg_{0.44}ReO_3$. We point out the possibility that Hg 6s electrons, which are not fully bonded in the metal-metal bond, enhance the electron-phonon coupling.

To summarize, we have discovered a hexagonal bronze superconductor Hg_xReO_3 with the transition temperature $T_c = 7.7$ K, which is the highest among known hexagonal bronzes. The crystal structure has a unique aspect that open tunnels accommodate $(Hg_2)^{2+}$ polycations, the rattling of which is one candidate as an origin of the superconducting instability.

This work was supported by Special Coordination Funds for Promoting Science and Technology, Promotion of Environmental Improvement for Independence of Young Researchers, and Grant-in-Aid for Scientific Research (C) (No. 20550135). K. O. is grateful for support from the Special Postdoctoral Researcher's Program of RIKEN.

[1] A. R. Sweedler, C. J. Raub, and B. T. Matthias, Phys. Lett. **15**, 108 (1965).

[2] A. R. Sweedler, J. K. Hulm, B. T. Matthias, and T. H. Geballe, Phys. Lett. **19**, 82 (1965).

[3] P. E. Biersted, T. A. Bither, and F. J. Darnell, Solid State Commun. **4**, 25 (1966).

[4] A. Magneli, Nature (London) **169**, 791 (1952).

[5] S. Raj *et al.*, Phys. Rev. B **77**, 245120 (2008).

[6] A. Hussain, R. Gruehn, and C. H. Ruscher, J. Alloys Compd. **246**, 51 (1997).

[7] L. Kihlborg and A. Hussain, Mater. Res. Bull. **14**, 667 (1979).

[8] P. Labbe, M. Goreaud, B. Raveau, and J. C. Monier, Acta Crystallogr. Sect. E **34**, 1433 (1978).

[9] K. Tatsumi, M. Hibino, and T. Kudo, Solid State Ionics **96**, 35 (1997).

[10] R. K. Stanley, R. C. Morris, and W. G. Moulton, Phys. Rev. B **20**, 1903 (1979).

[11] M. R. Skokan, W. G. Moulton, and R. C. Morris, Phys. Rev. B **20**, 3670 (1979).

[12] A. W. Sleight, T. A. Bither, and P. E. Biersted, Solid State Commun. **7**, 299 (1969).

[13] We also synthesized Hg_xReO_3 from a 1:1 molar ratio of HgO and Re, which results in considerable amounts of Re impurities. T_c of this sample is 8.0 K, which is slightly higher than $T_c = 7.7$ K of the sample grown from HgO and ReO_2. The fact that a reducing atmosphere leads to a higher T_c markedly contrasts with the case of tungsten bronzes, where an acid etching increases T_c [J. P. Remeika *et al.*, Phys. Lett. A **24**, 565 (1967).].

[14] H. Taniguchi *et al.*, J. Phys. Soc. Jpn. **72**, 468 (2003).

[15] F. Izumi and T. Ikeda, Mater. Sci. Forum **321–324**, 198 (2000).

[16] For example, M. S. Schriewerpottgen and W. Z. Jeitschko, Z. Anorg. Allg. Chem. **620**, 1855 (1994).

[17] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, Phys. Rev. **108**, 1175 (1957).

[18] L. C. Ting *et al.*, Chin. J. Phys. (Taipei) **45**, 237 (2007).

[19] F. C. Zumsteg and T. Pearsall, Solid State Commun. **16**, 751 (1975).

[20] W. A. Kamitakahara, K. Scharnberg, and H. R. Shanks, Phys. Rev. Lett. **43**, 1607 (1979).

[21] V. Keppens *et al.*, Nature (London) **395**, 876 (1998).

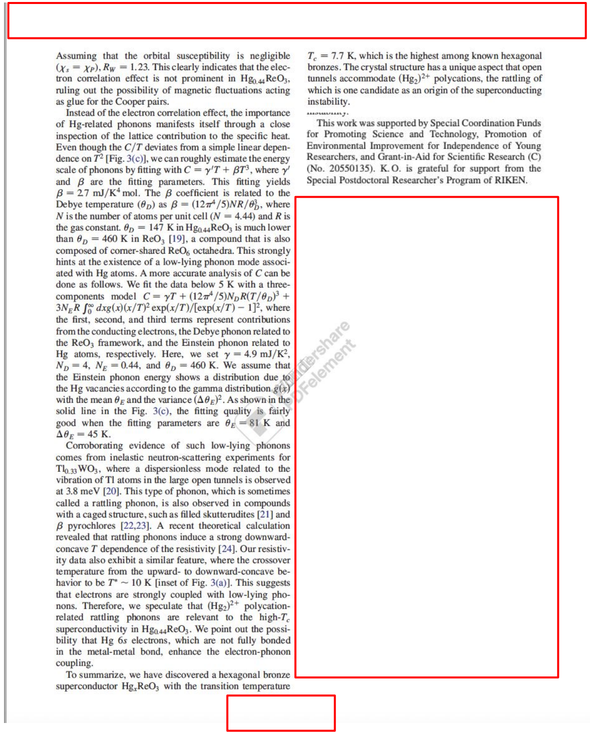
[22] K. Sasai *et al.*, J. Phys. Soc. Jpn. **76**, 104603 (2007).

[23] Y. Nagao *et al.*, J. Phys. Soc. Jpn. **78**, 064702 (2009).

[24] T. Dahm and K. Ueda, Phys. Rev. Lett. **99**, 187003 (2007).

017001-4

Before pre-processing



Assuming that the orbital susceptibility is negligible ($\chi_s = \chi_P$), $R_W = 1.23$. This clearly indicates that the electron correlation effect is not prominent in $Hg_{0.44}ReO_3$, ruling out the possibility of magnetic fluctuations acting as glue for the Cooper pairs.

Instead of the electron correlation effect, the importance of Hg-related phonons manifests itself through a close inspection of the lattice contribution to the specific heat. Even though the C/T deviates from a simple linear dependence on T^2 [Fig. 3(c)], we can roughly estimate the energy scale of phonons by fitting with $C = \gamma' T + \beta T^3$, where γ' and β are the fitting parameters. This fitting yields $\beta = 2.7$ mJ/K^4 mol. The β coefficient is related to the Debye temperature (θ_D) as $\beta = (12\pi^4/5)NR/\theta_D^3$, where N is the number of atoms per unit cell ($N = 4.44$) and R is the gas constant. $\theta_D = 147$ K in $Hg_{0.44}ReO_3$ is much lower than $\theta_D = 460$ K in ReO_3 [19], a compound that is also composed of corner-shared ReO_6 octahedra. This strongly hints at the existence of a low-lying phonon mode associated with Hg atoms. A more accurate analysis of C can be done as follows. We fit the data below 5 K with a three-components model $C = \gamma T + (12\pi^4/5)N_D R(T/\theta_D)^3 + 3N_E R \int_0^\infty dx g(x)(x/T)^2 \exp(x/T)/[\exp(x/T) - 1]^2$, where the first, second, and third terms represent contributions from the conducting electrons, the Debye phonon related to the ReO_3 framework, and the Einstein phonon related to Hg atoms, respectively. Here, we set $\gamma = 4.9$ mJ/K^2, $N_D = 4$, $N_E = 0.44$, and $\theta_D = 460$ K. We assume that the Einstein phonon energy shows a distribution due to the Hg vacancies according to the gamma distribution $g(x)$ with the mean θ_E and the variance $(\Delta\theta_E)^2$. As shown in the solid line in the Fig. 3(c), the fitting quality is fairly good when the fitting parameters are $\theta_E = 81$ K and $\Delta\theta_E = 45$ K.

Corroborating evidence of such low-lying phonons comes from inelastic neutron-scattering experiments for $Tl_{0.33}WO_3$, where a dispersionless mode related to the vibration of Tl atoms in the large open tunnels is observed at 3.8 meV [20]. This type of phonon, which is sometimes called a rattling phonon, is also observed in compounds with a caged structure, such as filled skutterudites [21] and β pyrochlores [22,23]. A recent theoretical calculation revealed that rattling phonons induce a strong downward-concave T dependence of the resistivity [24]. Our resistivity data also exhibit a similar feature, where the crossover temperature from the upward- to downward-concave behavior to be $T^* \sim 10$ K [inset of Fig. 3(a)]. This suggests that electrons are strongly coupled with low-lying phonons. Therefore, we speculate that $(Hg_2)^{2+}$ polycation-related rattling phonons are relevant to the high-T_c superconductivity in $Hg_{0.44}ReO_3$. We point out the possibility that Hg 6s electrons, which are not fully bonded in the metal-metal bond, enhance the electron-phonon coupling.

To summarize, we have discovered a hexagonal bronze superconductor Hg_xReO_3 with the transition temperature $T_c = 7.7$ K, which is the highest among known hexagonal bronzes. The crystal structure has a unique aspect that open tunnels accommodate $(Hg_2)^{2+}$ polycations, the rattling of which is one candidate as an origin of the superconducting instability.

This work was supported by Special Coordination Funds for Promoting Science and Technology, Promotion of Environmental Improvement for Independence of Young Researchers, and Grant-in-Aid for Scientific Research (C) (No. 20550135). K. O. is grateful for support from the Special Postdoctoral Researcher's Program of RIKEN.

After pre-processing

③ 前処理
Pre-processing

PDFelement

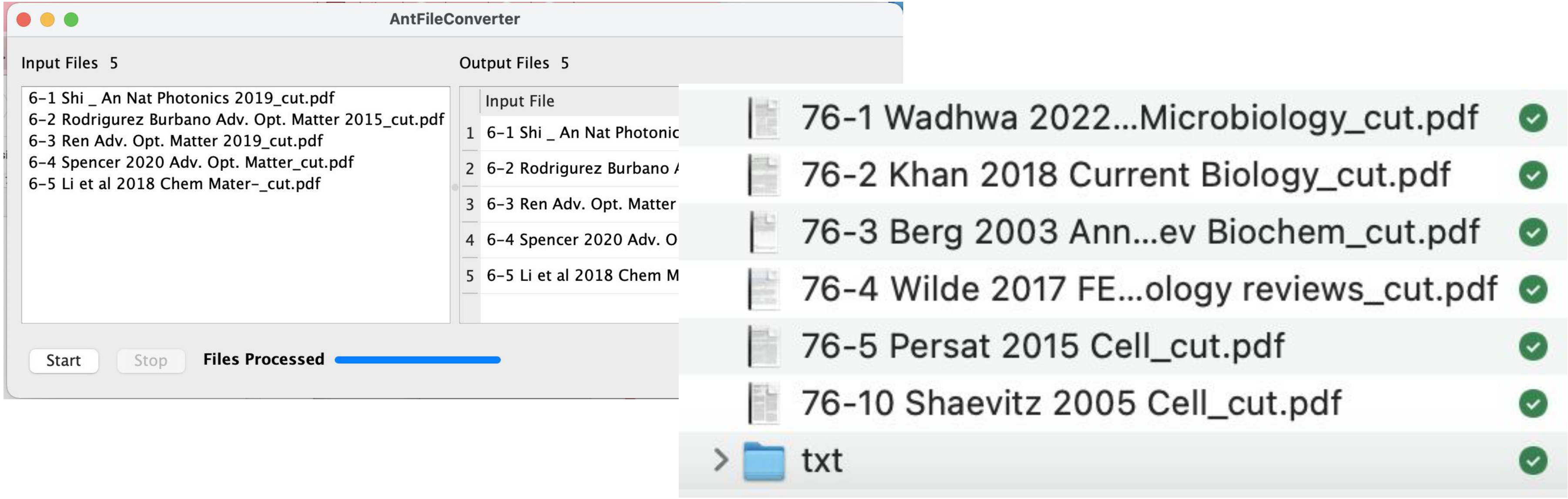
MATLAB

(63-116 Ohgushi)

Pre-Processing PDF-> txt

or MATLAB

- Simplest option **AntFileConverter (time: 15 mins for 330 files)**
- Programming option **MATLAB (time: 90 sec for 330 files)**



AntFileConverter

Input Files 5

6–1 Shi _ An Nat Photonics 2019_cut.pdf
6–2 Rodrigurez Burbano Adv. Opt. Matter 2015_cut.pdf
6–3 Ren Adv. Opt. Matter 2019_cut.pdf
6–4 Spencer 2020 Adv. Opt. Matter_cut.pdf
6–5 Li et al 2018 Chem Mater–_cut.pdf

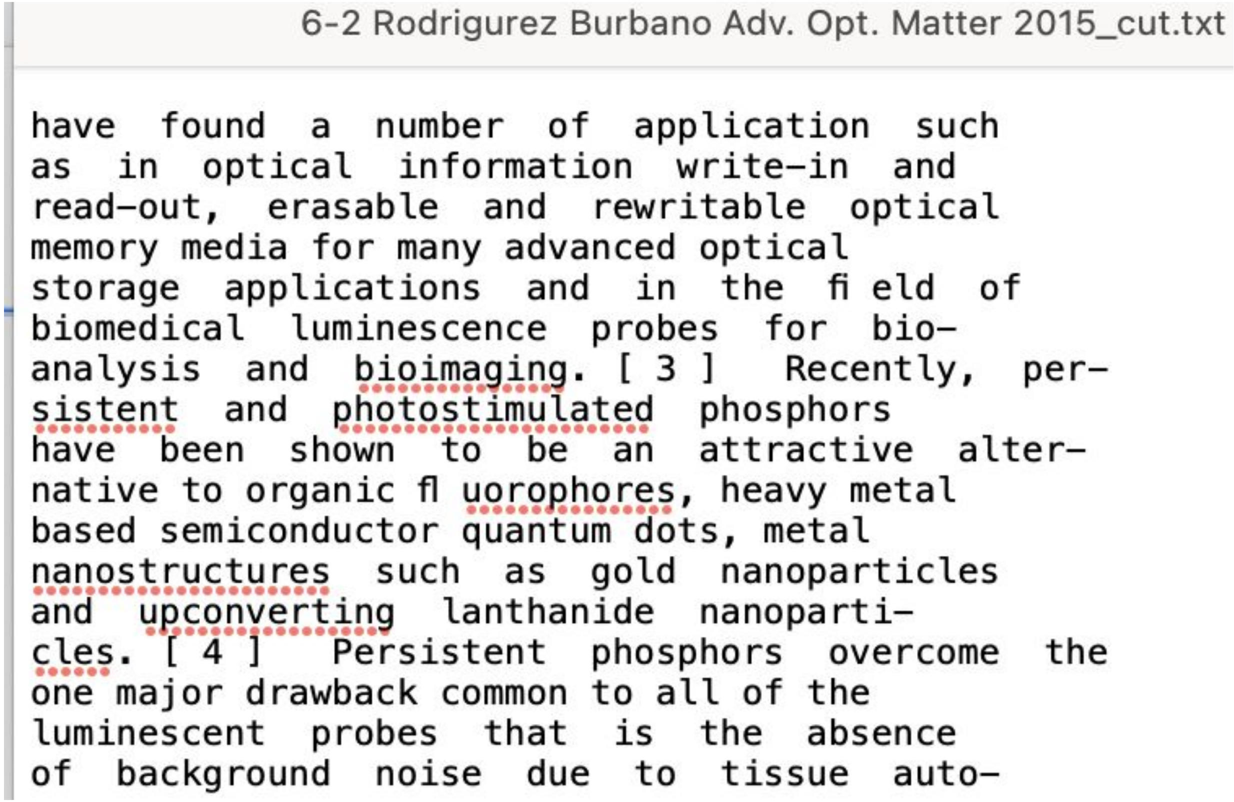
Output Files 5

	Input File
1	6–1 Shi _ An Nat Photonic
2	6–2 Rodrigurez Burbano A
3	6–3 Ren Adv. Opt. Matter
4	6–4 Spencer 2020 Adv. O
5	6–5 Li et al 2018 Chem M

76-1 Wadhwa 2022...Microbiology_cut.pdf ✓
76-2 Khan 2018 Current Biology_cut.pdf ✓
76-3 Berg 2003 Ann...ev Biochem_cut.pdf ✓
76-4 Wilde 2017 FE...ology reviews_cut.pdf ✓
76-5 Persat 2015 Cell_cut.pdf ✓
76-10 Shaevitz 2005 Cell_cut.pdf ✓
txt ✓

Sample Pre-processing (PDF-> txt file)

When creating a corpus and high frequency word list what problems can you identify in this txt file?

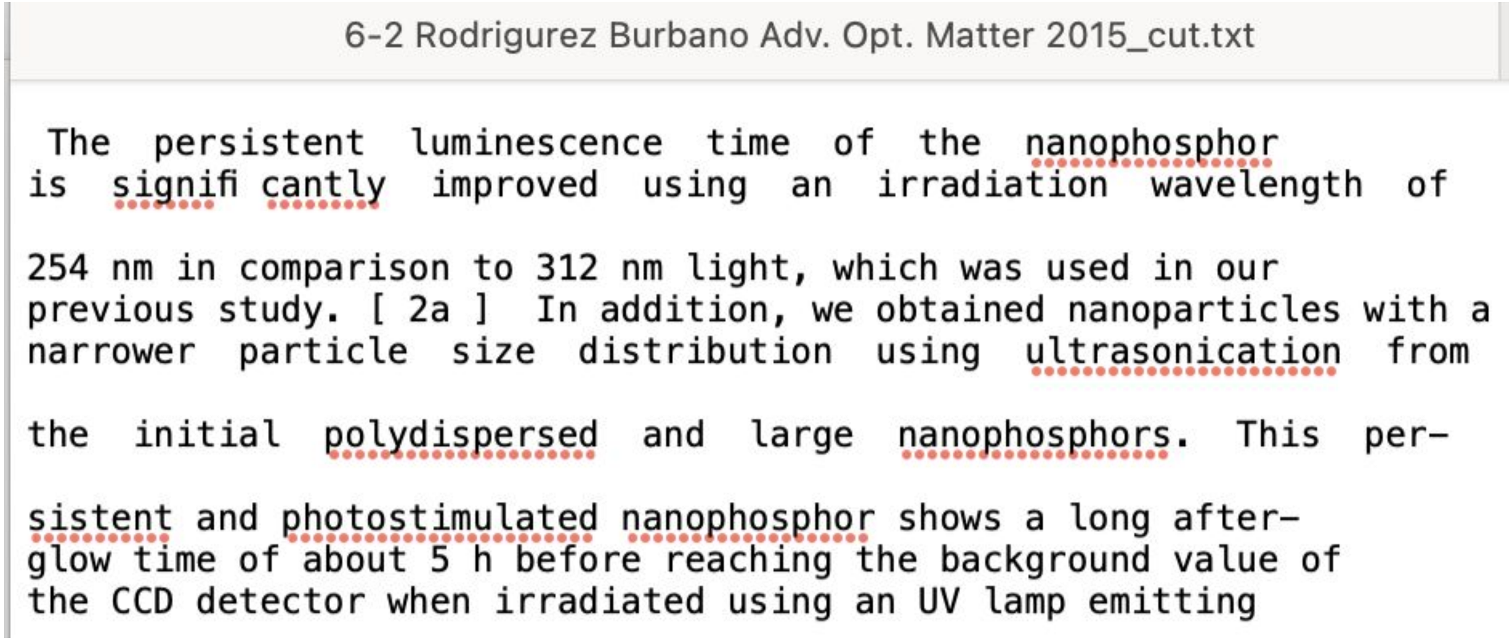


6-2 Rodrigurez Burbano Adv. Opt. Matter 2015_cut.txt

```
have  found  a  number  of  application  such
as  in  optical  information  write-in  and
read-out,  erasable  and  rewritable  optical
memory media for many advanced optical
storage  applications  and  in  the  fi eld  of
biomedical  luminescence  probes  for  bio-
analysis  and  bioimaging. [ 3 ]   Recently,  per-
sistent  and  photostimulated  phosphors
have  been  shown  to  be  an  attractive  alter-
native to organic fl uorophores, heavy metal
based semiconductor quantum dots, metal
nanostructures  such  as  gold  nanoparticles
and  upconverting  lanthanide  nanoparti-
cles. [ 4 ]   Persistent  phosphors  overcome  the
one major drawback common to all of the
luminescent  probes  that  is  the  absence
of  background  noise  due  to  tissue  auto-
```

Sample Pre-processing (PDF-> txt file)

When creating a corpus and high frequency word list what issues can you identify in this txt file?

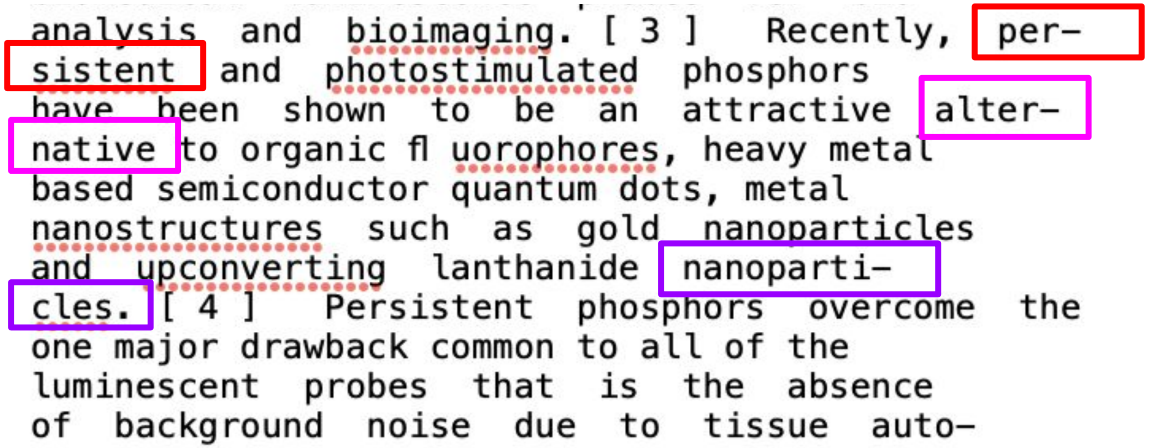


6-2 Rodrigurez Burbano Adv. Opt. Matter 2015_cut.txt

```
 The  persistent  luminescence  time  of  the  nanophosphor
is  signifi cantly  improved  using  an  irradiation  wavelength  of

254 nm in comparison to 312 nm light, which was used in our
previous study. [ 2a ]  In addition, we obtained nanoparticles with a
narrower  particle  size  distribution  using  ultrasonication  from

the  initial  polydispersed  and  large  nanophosphors.  This  per–

sistent and photostimulated nanophosphor shows a long after–
glow time of about 5 h before reaching the background value of
the CCD detector when irradiated using an UV lamp emitting
```

Some Issues in Pre-Processing

- Hypenation
  (e.g. persistent vs per-sistent)



analysis and bioimaging. [3] Recently, per–
sistent and photostimulated phosphors
have been shown to be an attractive alter–
native to organic fl uorophores, heavy metal
based semiconductor quantum dots, metal
nanostructures such as gold nanoparticles
and upconverting lanthanide nanoparti–
cles. [4] Persistent phosphors overcome the
one major drawback common to all of the
luminescent probes that is the absence
of background noise due to tissue auto–

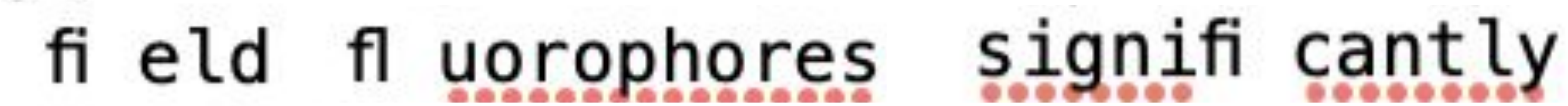
per-sistent

alter-native

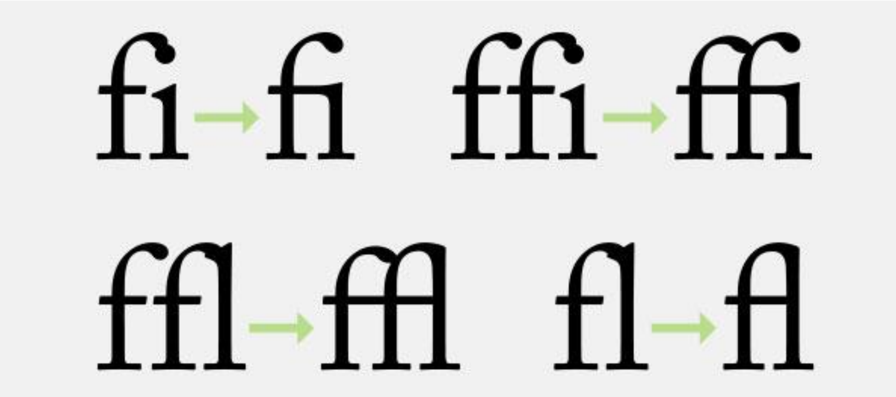
nanoparti-cles

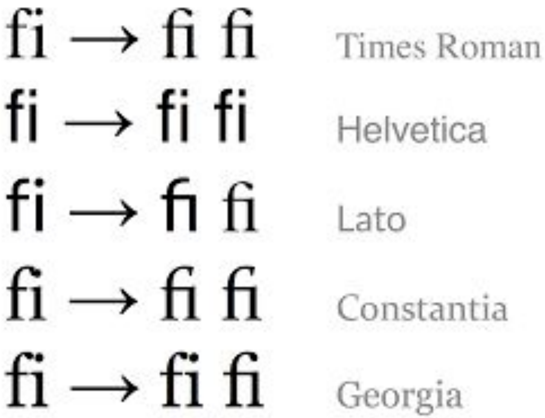
Some Issues in Pre-Processing

- Some fonts are problematic. The file conversion tool may see an image not two or three different sequences of letters.



ﬁ eld ﬂ uorophores signiﬁ cantly





③ Pre-Processing

Using MATLAB “Text Analytics Toolbox”, PDF files were converted to text files.
The program successfully processes pdf to create a cleaner txt file.

Basic Program

- Converts pdf to txt

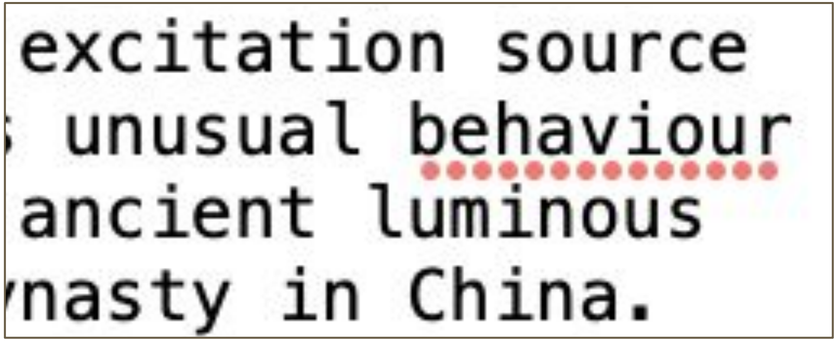
Additional functions

- Removes hyphens to make one word.
- Removes spaces after fi and fl to make one word.
- Removes 1-2 letter “words”.
- Removes all number words (e.g. one, two, three…thousand)

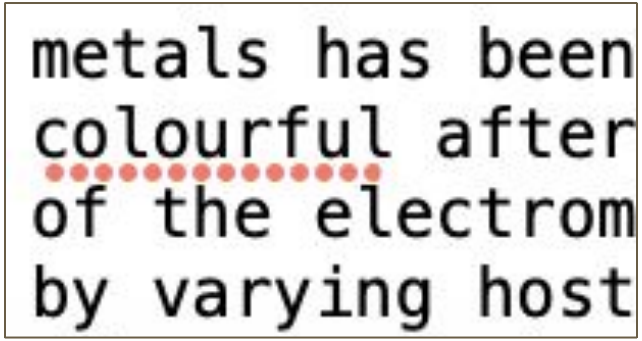
③ 前処理
Pre-processing
PDFelement
MATLAB

Pending issues: British and American Spelling

How does the program count behavior and behaviour?



excitation source
unusual behaviour
ancient luminous
nasty in China.

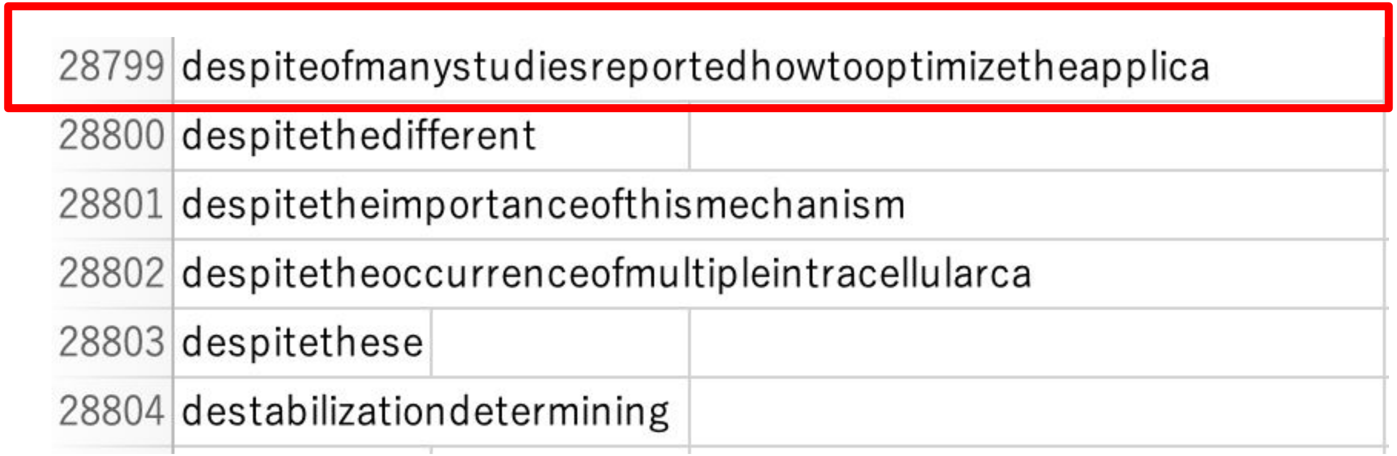


metals has been
colourful after
of the electrom
by varying host

(6-1 Shi & An)

Pending issues: Accuracy of Text Extraction

This is an example of the 28799th high frequency word. What is the issue?



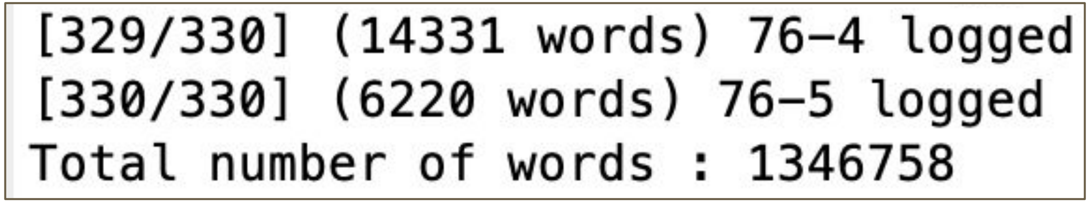
28799	despiteofmanystudiesreportedhowtooptimizetheapplica
28800	despitethedifferent
28801	despitetheimportanceofthismechanism
28802	despitetheoccurrenceofmultipleintracellularca
28803	despitethese
28804	destabilizationdetermining

Pending Issues: Token Count

How to report the correct number of tokens (words)

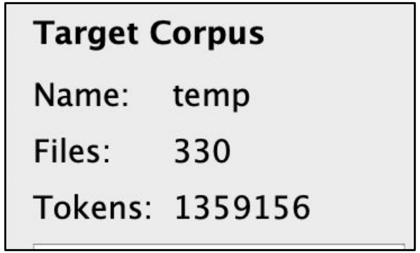
MATLAB (1346758 tokens)
AntConc (1359156 tokens) ➡

10% difference
(12,398 tokens)



```
[329/330] (14331 words) 76-4 logged
[330/330] (6220 words) 76-5 logged
Total number of words : 1346758
```

MATLAB



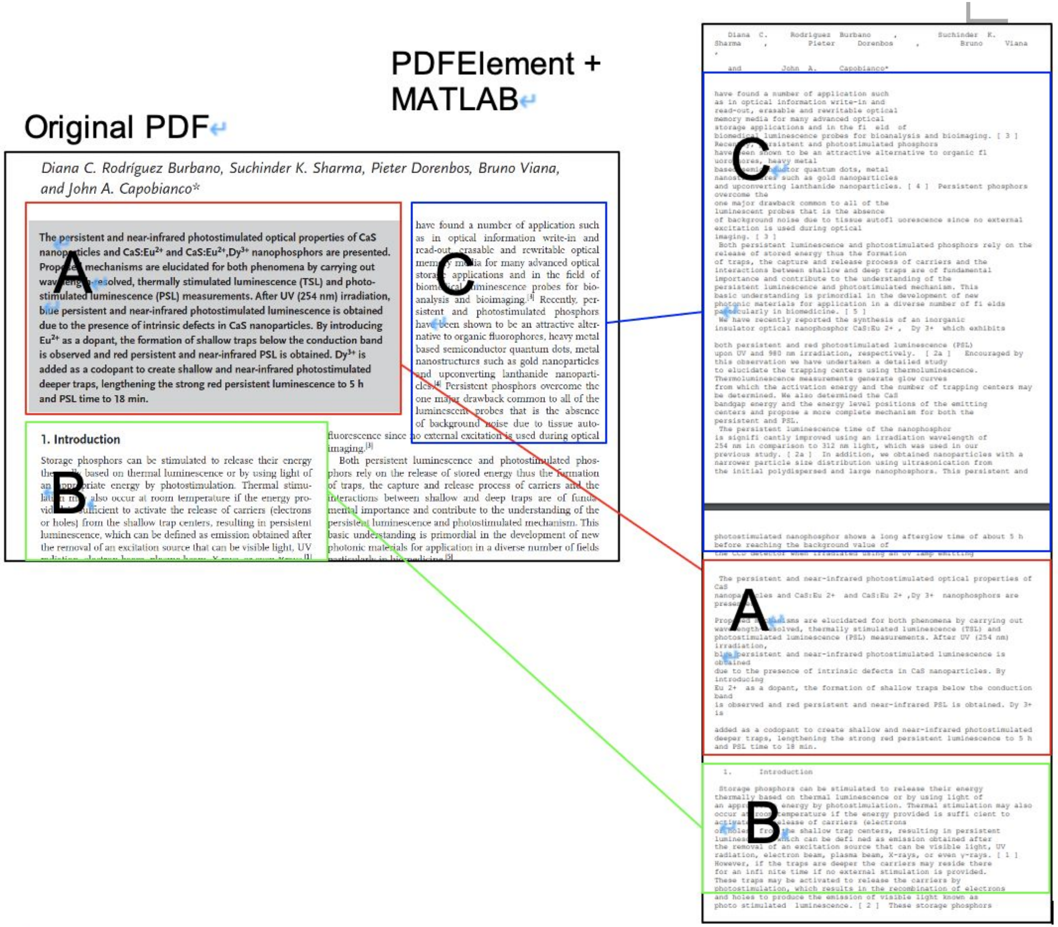
Target Corpus

Name: temp
Files: 330
Tokens: 1359156

AntConc

Pending issues: Corpus

- Order of text is jumbled up.
- For the purpose of this study (creating a high frequency word list), the order does not matter.
- For a clean corpus, future aims would be to compare edit the text document further.


Original PDF
PDFElement + MATLAB
A
B
C
C
A
B

*Diana C. Rodríguez Burbano, Suchinder K. Sharma, Pieter Dorenbos, Bruno Viana, and John A. Capobianco**

The persistent and near-infrared photostimulated optical properties of CaS nanoparticles and CaS:Eu^{2+} and CaS:Eu^{2+},Dy^{3+} nanophosphors are presented. Proposed mechanisms are elucidated for both phenomena by carrying out wavelength-resolved, thermally stimulated luminescence (TSL) and photostimulated luminescence (PSL) measurements. After UV (254 nm) irradiation, blue persistent and near-infrared photostimulated luminescence is obtained due to the presence of intrinsic defects in CaS nanoparticles. By introducing Eu^{2+} as a dopant, the formation of shallow traps below the conduction band is observed and red persistent and near-infrared PSL is obtained. Dy^{3+} is added as a codopant to create shallow and near-infrared photostimulated deeper traps, lengthening the strong red persistent luminescence to 5 h and PSL time to 18 min.

1. Introduction

Storage phosphors can be stimulated to release their energy thermally based on thermal luminescence or by using light of an appropriate energy by photostimulation. Thermal stimulation may also occur at room temperature if the energy provided is sufficient to activate the release of carriers (electrons or holes) from the shallow trap centers, resulting in persistent luminescence, which can be defined as emission obtained after the removal of an excitation source that can be visible light, UV radiation, electron beam, plasma beam, X-rays, or even γ-rays.[1]

have found a number of application such as in optical information write-in and read-out, erasable and rewritable optical memory media for many advanced optical storage applications and in the field of biomedical luminescence probes for bioanalysis and bioimaging.[3] Recently, persistent and photostimulated phosphors have been shown to be an attractive alternative to organic fluorophores, heavy metal based semiconductor quantum dots, metal nanostructures such as gold nanoparticles and upconverting lanthanide nanoparticles.[4] Persistent phosphors overcome the one major drawback common to all of the luminescent probes that is the absence of background noise due to tissue autofluorescence since no external excitation is used during optical imaging.[3]

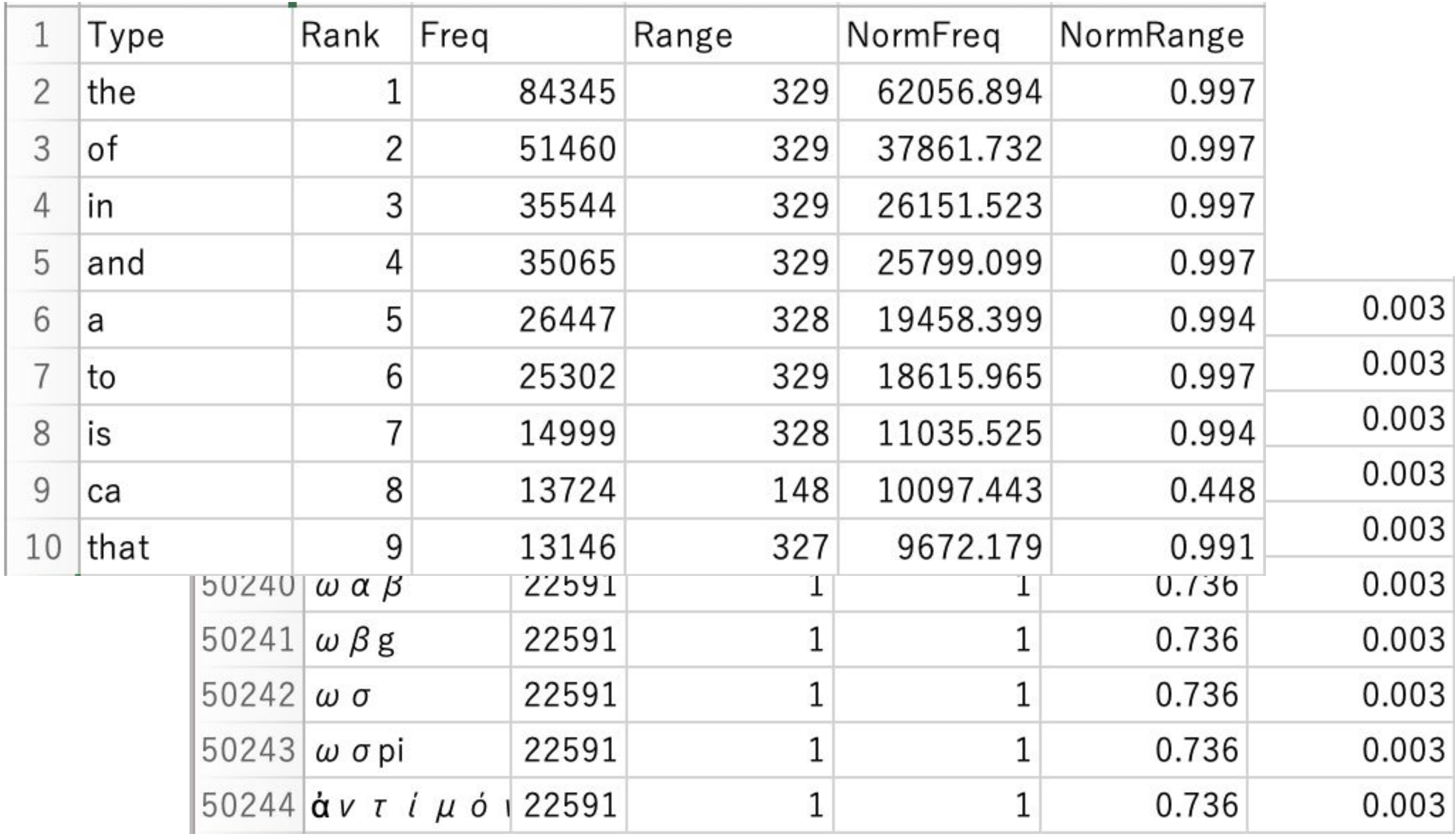
Both persistent luminescence and photostimulated phosphors rely on the release of stored energy thus the formation of traps, the capture and release process of carriers and the interactions between shallow and deep traps are of fundamental importance and contribute to the understanding of the persistent luminescence and photostimulated mechanism. This basic understanding is primordial in the development of new photonic materials for application in a diverse number of fields particularly in biomedicine.[5]

```
   Diana  C.    Rodriguez  Burbano    ,        Suchinder  K.
Sharma    ,         Pieter    Dorenbos     ,        Bruno     Viana
,

   and        John  A.    Capobianco*

have found a number of application such
as in optical information write-in and
read-out, erasable and rewritable optical
memory media for many advanced optical
storage applications and in the fi  eld  of
biomedical luminescence probes for bioanalysis and bioimaging. [ 3 ]
Recently, persistent and photostimulated phosphors
have been shown to be an attractive alternative to organic fl
uorophores, heavy metal
based semiconductor quantum dots, metal
nanostructures such as gold nanoparticles
and upconverting lanthanide nanoparticles. [ 4 ]  Persistent phosphors
overcome the
one major drawback common to all of the
luminescent probes that is the absence
of background noise due to tissue autofl uorescence since no external
excitation is used during optical
imaging. [ 3 ]
 Both persistent luminescence and photostimulated phosphors rely on the
release of stored energy thus the formation
of traps, the capture and release process of carriers and the
interactions between shallow and deep traps are of fundamental
importance and contribute to the understanding of the
persistent luminescence and photostimulated mechanism. This
basic understanding is primordial in the development of new
photonic materials for application in a diverse number of fi elds
particularly in biomedicine. [ 5 ]
 We have recently reported the synthesis of an inorganic
insulator optical nanophosphor CaS:Eu 2+ ,  Dy 3+  which exhibits

both persistent and red photostimulated luminescence (PSL)
upon UV and 980 nm irradiation, respectively.  [ 2a ]   Encouraged by
this observation we have undertaken a detailed study
to elucidate the trapping centers using thermoluminescence.
Thermoluminescence measurements generate glow curves
from which the activation energy and the number of trapping centers may
be determined. We also determined the CaS
bandgap energy and the energy level positions of the emitting
centers and propose a more complete mechanism for both the
persistent and PSL.
 The persistent luminescence time of the nanophosphor
is signifi cantly improved using an irradiation wavelength of
254 nm in comparison to 312 nm light, which was used in our
previous study. [ 2a ]  In addition, we obtained nanoparticles with a
narrower particle size distribution using ultrasonication from
the initial polydispersed and large nanophosphors. This persistent and


photostimulated nanophosphor shows a long afterglow time of about 5 h
before reaching the background value of
the CCD detector when irradiated using an UV lamp emitting

 The persistent and near-infrared photostimulated optical properties of
CaS
nanoparticles and CaS:Eu 2+  and CaS:Eu 2+ ,Dy 3+  nanophosphors are
presented.

Proposed mechanisms are elucidated for both phenomena by carrying out
wavelength-resolved, thermally stimulated luminescence (TSL) and
photostimulated luminescence (PSL) measurements. After UV (254 nm)
irradiation,
blue persistent and near-infrared photostimulated luminescence is
obtained
due to the presence of intrinsic defects in CaS nanoparticles. By
introducing
Eu 2+  as a dopant, the formation of shallow traps below the conduction
band
is observed and red persistent and near-infrared PSL is obtained. Dy 3+
is

added as a codopant to create shallow and near-infrared photostimulated
deeper traps, lengthening the strong red persistent luminescence to 5 h
and PSL time to 18 min.

  1.       Introduction

 Storage phosphors can be stimulated to release their energy
thermally based on thermal luminescence or by using light of
an appropriate energy by photostimulation. Thermal stimulation may also
occur at room temperature if the energy provided is suffi cient to
activate the release of carriers (electrons
or holes) from the shallow trap centers, resulting in persistent
luminescence, which can be defi ned as emission obtained after
the removal of an excitation source that can be visible light, UV
radiation, electron beam, plasma beam, X-rays, or even γ-rays. [ 1 ]
However, if the traps are deeper the carriers may reside there
for an infi nite time if no external stimulation is provided.
These traps may be activated to release the carriers by
photostimulation, which results in the recombination of electrons
and holes to produce the emission of visible light known as
photo stimulated  luminescence. [ 2 ]  These storage phosphors
```

④ Data Processing

MATLAB processed txt was run in AntConc to create an initial high-frequency word list.

④ Data Processing



1	Type	Rank	Freq	Range	NormFreq	NormRange
2	the	1	84345	329	62056.894	0.997
3	of	2	51460	329	37861.732	0.997
4	in	3	35544	329	26151.523	0.997
5	and	4	35065	329	25799.099	0.997
6	a	5	26447	328	19458.399	0.994
7	to	6	25302	329	18615.965	0.997
8	is	7	14999	328	11035.525	0.994
9	ca	8	13724	148	10097.443	0.448
10	that	9	13146	327	9672.179	0.991

						0.003
						0.003
						0.003
						0.003
						0.003
50240	ω α β	22591	1	1	0.736	0.003
50241	ω β g	22591	1	1	0.736	0.003
50242	ω σ	22591	1	1	0.736	0.003
50243	ω σ pi	22591	1	1	0.736	0.003
50244	ἀν τ ί μ ό ι	22591	1	1	0.736	0.003

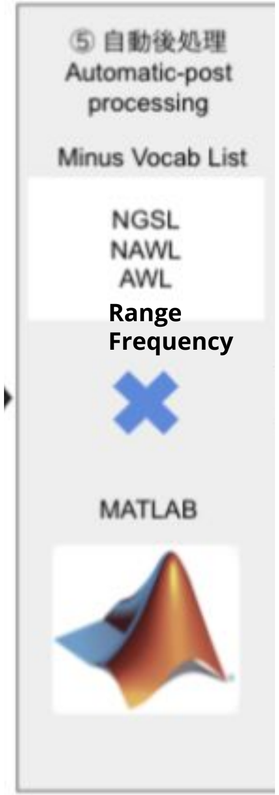
④ 処理
Data Processing

Antconc

⑤ Automatic Post-Processing

Additional functions with MATLAB (all adjustable)

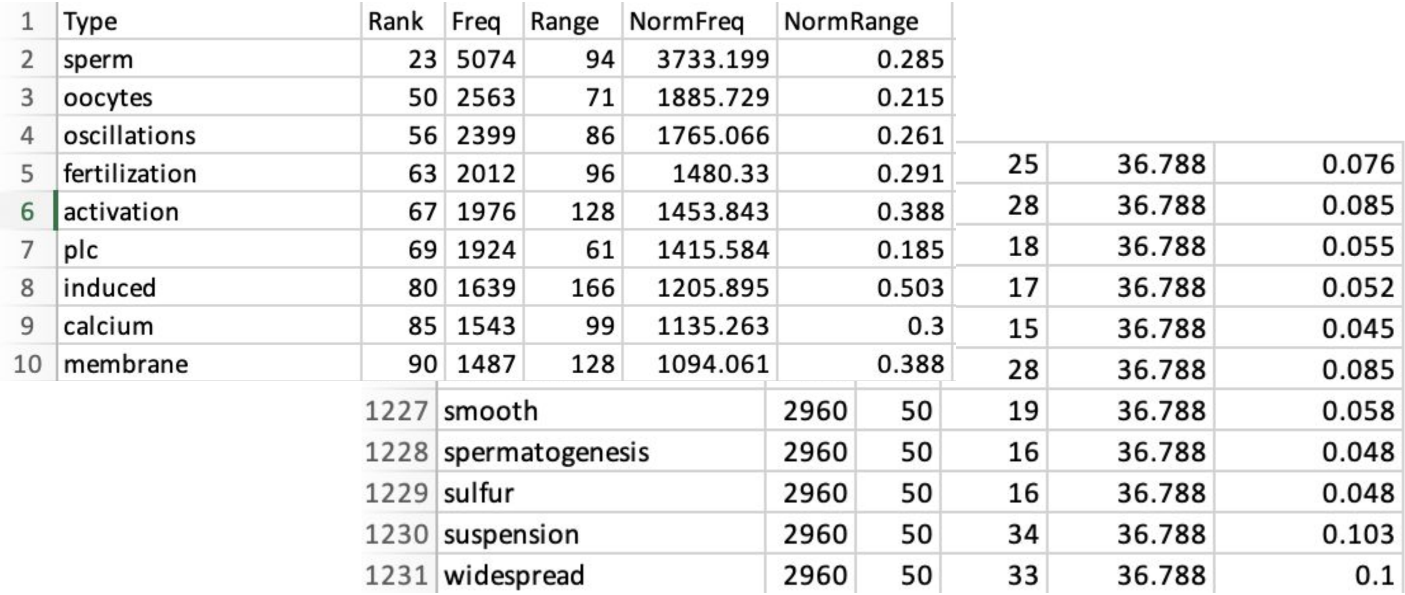
- -Removed 1st to 2nd 2000 NGSL
- -Removed supplementary data in NGSL
- -AWL not removed
- -Range 15 (out of 330)
- -Frequency 50 (out of 330)



⑤ Automatic Post-Processing

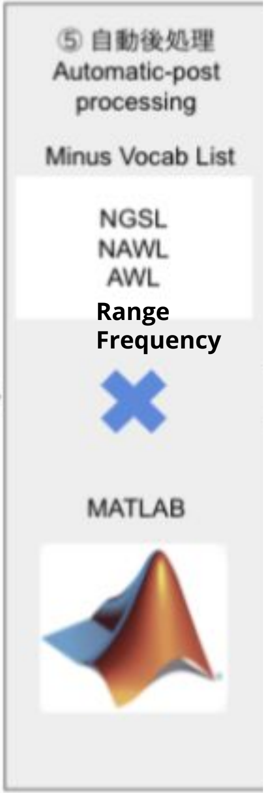
New List: 1231 words (Range >3, Frequency >50)

*188 out of 330 papers from a professor whose research is on fertilization.



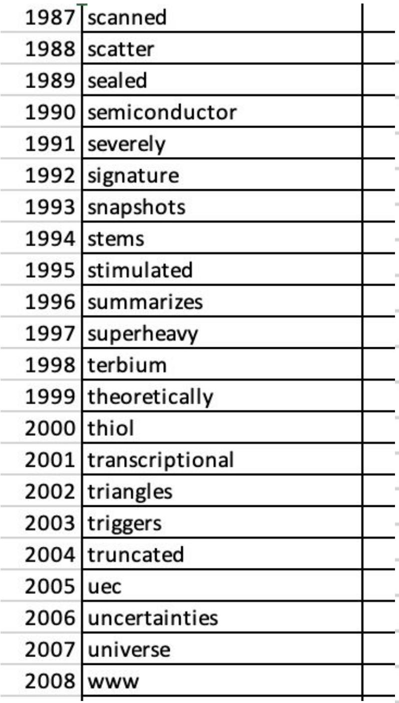
1	Type	Rank	Freq	Range	NormFreq	NormRange
2	sperm	23	5074	94	3733.199	0.285
3	oocytes	50	2563	71	1885.729	0.215
4	oscillations	56	2399	86	1765.066	0.261
5	fertilization	63	2012	96	1480.33	0.291
6	activation	67	1976	128	1453.843	0.388
7	plc	69	1924	61	1415.584	0.185
8	induced	80	1639	166	1205.895	0.503
9	calcium	85	1543	99	1135.263	0.3
10	membrane	90	1487	128	1094.061	0.388

				25	36.788	0.076
				28	36.788	0.085
				18	36.788	0.055
				17	36.788	0.052
				15	36.788	0.045
				28	36.788	0.085
1227	smooth	2960	50	19	36.788	0.058
1228	spermatogenesis	2960	50	16	36.788	0.048
1229	sulfur	2960	50	16	36.788	0.048
1230	suspension	2960	50	34	36.788	0.103
1231	widespread	2960	50	33	36.788	0.1

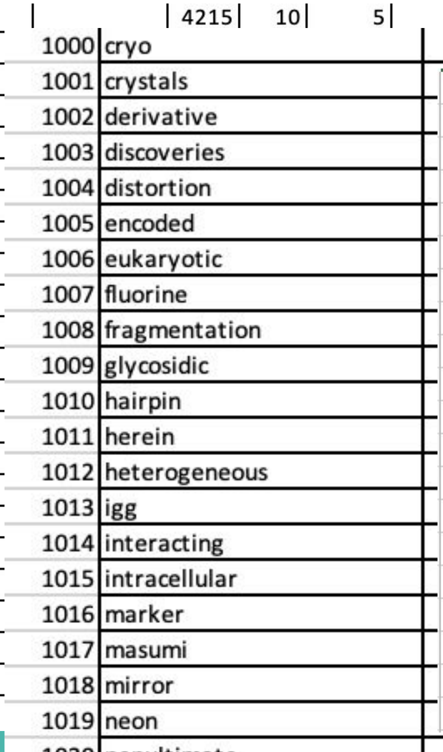


Result: Sample List from 174 files processed in March

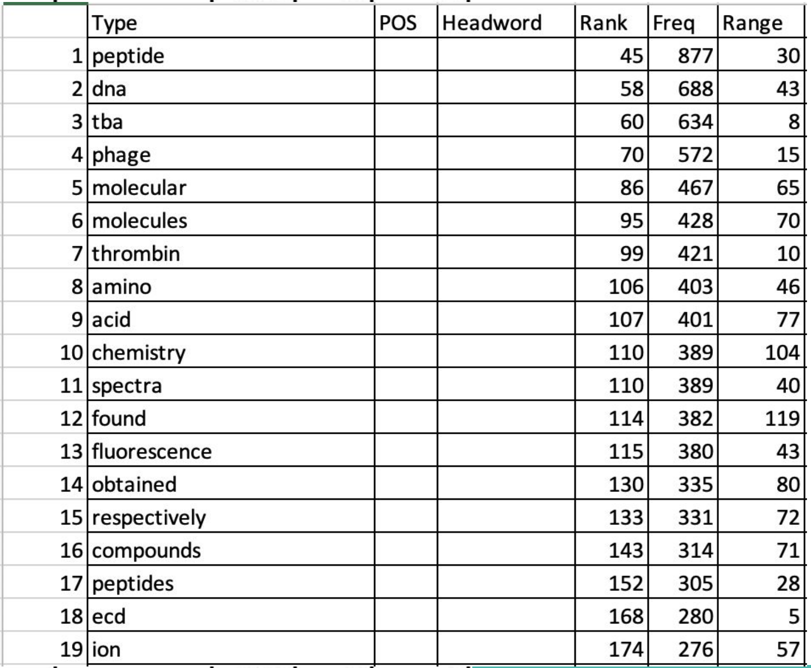
List varies with the selected corpus



1987	scanned
1988	scatter
1989	sealed
1990	semiconductor
1991	severely
1992	signature
1993	snapshots
1994	stems
1995	stimulated
1996	summarizes
1997	superheavy
1998	terbium
1999	theoretically
2000	thiol
2001	transcriptional
2002	triangles
2003	triggers
2004	truncated
2005	uec
2006	uncertainties
2007	universe
2008	www



			4215	10	5
1000	cryo		2474	22	5
1001	crystals				
1002	derivative				
1003	discoveries				
1004	distortion				
1005	encoded				
1006	eukaryotic				
1007	fluorine				
1008	fragmentation				
1009	glycosidic				
1010	hairpin				
1011	herein				
1012	heterogeneous				
1013	igg				
1014	interacting				
1015	intracellular				
1016	marker				
1017	masumi				
1018	mirror				
1019	neon				



	Type	POS	Headword	Rank	Freq	Range
1	peptide			45	877	30
2	dna			58	688	43
3	tba			60	634	8
4	phage			70	572	15
5	molecular			86	467	65
6	molecules			95	428	70
7	thrombin			99	421	10
8	amino			106	403	46
9	acid			107	401	77
10	chemistry			110	389	104
11	spectra			110	389	40
12	found			114	382	119
13	fluorescence			115	380	43
14	obtained			130	335	80
15	respectively			133	331	72
16	compounds			143	314	71
17	peptides			152	305	28
18	ecd			168	280	5
19	ion			174	276	57

⑥ Manual Post-Processing

A science major and science professor in the field of biochemistry is being consulted for words that can be removed from the list.

⑥手動後処理
Manual-post processing
Science Professor Check
High Frequency Word List

Discussion

Current corpus (1.3 million tokens) from Chemistry and Biotechnology program processed well

Used MATLAB program to make the corpus and high frequency word list development more efficient

Work required to carefully clean the corpus and generate high-frequency word lists

Research lab specific data should be used to generate high frequency word lists

Future Work

Compile a larger corpus that covers all four programs with recommendations from the different programs in equal ratio (Coxhead & Hirsch 2000)

Further considerations of range, frequency and dispersion required (Coxhead & Hirsch, 2000)

Acknowledgements

Many thanks to Hibiya for his MATLAB programming skills, and to the corpus pre-processing team; Edgarito, Tatsuki, Takashi, Mitsuki, Rina, Li

This project was partially funded by the University of Electro-Communications

MATLAB Programmer Hibiya

References

Biber, D., Conrad, S., & reppen, R. (1998). *Corpus linguistics: Investigating language structure and use.* Cambridge University Press.

Coxhead, A. (2000). A new academic word list. TESOL Quarterly, 34(2), 213–238. https://doi.org/10.2307/3587951.

Coxhead, A. & Hirsch, D. (2007). A pilot science-specific word list. *Rev. franç. de linguistique appliquée*, XII-2, 65–78.

Dang, T.N.Y. (2020). Corpus-based word lists in second language vocabulary research, learning, and teaching. In S. Webb (ed.), The Routledge handbook of vocabulary studies. Routledge.

Gilner, L. (2011). A primer on the General Service List. *Reading in a Foreign Language, 23*, 65–83.

Kelly, P. (1991). Lexical ignorance: The main obstacle to listening comprehension with advanced foreign language learners. *IRAL, 29*(2), 135–149.

He, X., & Godfroid, A. (2019). Choosing words to teach: A novel method for vocabulary selection and its practical application. *TESOL Quarterly, 53*, 348–371.

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics, 39*(6), 823–845. https://doi.org/10.1093/applin/amw050

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*(1), 59–82.

Nation, I. S. P. (2016). Making word lists. In The Vocab@Tokyo 2016 Handbook (p. 34). Retrieved from http://vli-journal.org/wp/vocab-at-tokyo-handbook-2016/

Nesi, H. (2013). 21 ESP and Corpus Studies. *The handbook of English for specific purposes*, 407.

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes, 28*, 170–182.

Webb, S., & Macalister, J. (2012). Is text written for children appropriate for L2 extensive reading? *TESOL Quarterly, 47*(2), 300–322.

Webb, S., & Rodgers, M. P. H. (2009). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407–427.

Xue, G., & Nation, I. S. P. (1984). A university word list. *Language Learning and Communication, 3*, 215–229.

Abstract

This is a report on a work-in-progress of the development of an in-house corpus and high-frequency word list of the corpus for a Science and Engineering university in Japan. Students read scientific articles as part of their required English courses. However, in an informal survey, while some students were positive about the prospects of reading specialized academic articles in English, others felt that it would be too challenging. In order to bridge the difficulty gap, an in-house corpus of articles recommended by the science faculty and a high-frequency word list of the corpus are being developed. Interviews and surveys will be conducted with selected members of the science department to understand the nature of articles written in English that these members would recommend for graduate students. The articles will be gathered to create a corpus of one million words, and processed for high-frequency words using AntConc (Version 4.0.2) (Anthony, 2021) a free online vocabulary profiling software. These will be compared against the new academic word list and further analysed for specialized words. The findings will help to construct an informative vocabulary list for the students in graduate school, and in the future, this could be further refined for undergraduate students.

Presentation

Developing an in-house corpus and high-frequency word list for science majors

Suwako Uehara
uehara.suwako@uec.ac.jp

July 8-10, 2022 at PanSIG
University of Nagano
https://pansig2022.edzil.la/session/2854
Schedule: 8th July at 18:25

PanSIG
University of Nagano
July 8-10, 2022