Length of an Abstract and Amount of Information

By KENZO HIRAYAMA
Research Laboratory, Fuji Photo Film, Co., Ltd., Minamiashigara, Odawara, Japan
Received June 10, 1963

Much has already been said about the tremendous increase in the technical literature. *Chemical Abstracts* printed 166,749 abstracts on 16,465 pages in 1962. Since the abstracts in *Chemical Abstracts* are printed in 6-point type, and each page contains two 83-line columns about 79 mm. wide, if all the abstracts for 1962 were printed in one line they would reach a length of about 220 km., about 25 times the height of Mt. Everest or 600 times that of the Empire State Building. Reading one of the 73 sections of *Chemical Abstracts*, such as the section on Analytical Chemistry, for one year is the equivalent of reading sentences running from the ground to top of the Empire State Building 17 times. When the amount of literature becomes this great, it seems timely to examine the relationship between the length of an abstract and amount of information it contains.

1. EXAMPLES OF ABSTRACTS

The abstracts used as material for this examination were based on the original Japanese paper shown in Fig. 1. (The tables and figures are omitted.) Five abstracts of this paper were found, written in Japanese, English, and German. These five abstracts, prepared by different abstractors and therefore having different aspects, were translated or rewritten into English by the same person. Preparation followed the route shown in Fig. 2.

Absorption Spectra and Chemical Structure. II. Solvent Effect

It was pointed out in part I that the λ_{max} of polyenes containing conjugated homochromophoric groups and auxochromic substituents may be expressed by

$$(\lambda_{max})^2 = A - BC^N \qquad (1)$$

$$= a + B(1 - C^N) \qquad (2)$$

where N is the homoconjugation index, which is determined by substituent and structural effects as well as by the number of conjugated chromophoric groups.

Thus, the following equation was given for alcholic solutions of polyene derivatives with auxochromic substituents

$$(\lambda^{alc}_{max})^2 = (36.98 - 39.10 \times 0.920^N) \times 10^4 \text{ m}\mu^2 \qquad (3)$$

$$= \{-2.12 + 39.10 \times (1 - 0.920^N)\} \times 10^4 \text{ m}\mu^2 \qquad (4)$$

We shall discuss the influence of the solvent on λ_{max}, *i.e.*, which parameters are effected by a change in solvent.

The parameters, (a, B and C) for carbon disulfide solutions of carotenoids were computed from the observed wave lengths; a and C are the same for alcohol and carbon disulfide solutions (-2.12 and 0.920, respectively), and only B is characteristic of the solvent. Formulas for the calculation of λ_{max} in various solvents were determined from the observed values with $a = -2.12$ and $c = 0.920$.

Solvent	(λmax), $\text{m}\mu^2 \times 10^{-4}$	
Methanol	$(36.30 - 38.42 \times 0.920^N)$	(5)
Hexane	$(36.62 - 38.72 \times 0.920^N)$	(6)
Petroleum ether	$(36.70 - 38.82 \times 0.920^N)$	(7)
Light ligroin	$(37.05 - 39.17 \times 0.920^N)$	(8)
Chloroform	$(38.60 - 40.72 \times 0.920^N)$	(9)
Benzene	$(39.20 - 41.32 \times 0.920^N)$	(10)
Carbon disulfide	$(42.06 - 44.18 \times 0.920^N)$	(11)

The calculated values of λ_{max} in Table I are based on these formulas, which differ only with respect to B which is characteristic for each solvent; the agreement between the calculated and observed values is good except for compounds 14 in carbon disulfide, 25 in chloroform and in benzene and 27 in hexane. In the case of rhodoviolascin 26, it was assumed that $N_{methoxy} = 0$; the agreement between the observed values and those calculated with $N_{methoxy} = 0$ indicates that this substituent exerts practically no bathochromic effect.

Figure 1 shows that there is an approximately linear relationship between B^{solv} and the refractive index of the solvent.[2] The values for B^{solv} in formulas 5–11 agree well with the values calculated by the equation (see Table II)

$$B^{solv}_{calcd} = 12.976 + 19.019 \times n^{20}\text{D} \qquad (12)$$

Because the number of observations made in ether, cyclohexane and pyridine was insufficient for the determination of B^{solv}, B^{solv}_{calcd} computed from equation 12 was used in the calculation of λ_{max} in these solvents (Table III); the good agreement between the calculated and experimental values indicates the reliability of relation 12.

Figure 1.

```
                 Author's          Translated
Author's      ↗  abstract in -->   English          ---> Abstract 1
original         Japanese          abstract by E
paper in      ↘
Japanese         Japanese          Translated
   |             abstract by  ->   English          ---> Abstract 2
   |             Abstractor A      abstract by E
   |
   |             English           English
   ↓             abstract by  ->   abstract          --> Abstract 3
English       ↗  Abstractor B      rewritten by E
translation      English           English
of the      ---> abstract by  ->   abstract          --> Abstract 4
original         Abstractor C      rewritten by E
paper         ↘
                 German            English
                 abstract by  ->   abstract           -> Abstract 5
                 Abstractor D      translated by E
```

Figure 2.

Abstract 1.—Absorption max. of polyenes possessing auxochrome can be expressed by $(\lambda_{max.})^2 = A - BC^N$ and $\lambda_{max.}$ in various solvents can be calcd. from this formula by varying B. Approx. linear relation was found to exist between B and n of the solvent.

Abstract 2.—Absorption max. of 57 kinds of polyene compds., including $H—(CH=CH)_3—H$, were measured in soln. of MeOH, hexane, petr. ether, ligroine, EtOH, $CHCl_3$, C_6H_6, and CS_2. A formula common to all these solvents can be obtained by changing B in the formula $(\lambda_{max.})^2 = A + BC^N$ for relationship between $\lambda_{max.}$ and structure. There is a linear relationship between the value of B and n of various solvents.

Abstract 3.[1]—The difference of $(\lambda_{max.})^2$, which increase as N increases, between a certain solvent and EtOH is shown to be the solvent effect of B and B has a linear relation to n_D^{20} of the solvent. Calcd. values of $\lambda_{max.}$ for 57 compds. in MeOH, hexane, Et_2O, EtOH, cyclohexane, $CHCl_3$, C_6H_6, pyridine, CS_2, petr. ether, or gasoline showed good agreement with observed values when $B_{calcd.}^{solv.} = 12.986 + 19.019\, n_D^{20}$.

Abstract 4.—The formula for calcg. the 1st absorption max. of polyenes can be expressed by $(\lambda_{max.})^2 = A - BC^N$. It is shown that the solvent modifies B and that there is a linear relationship between B and n_D^{20} of the solvent. The calcd. and observed 1st absorption max. are tabulated for 57 polyenes. The n_D^{20} and $B^{solv.}$ values were: MeOH, 1.3288 and 38.42; hexane, 1.3751 and 38.74; petr. ether, 1.37 and 38.82; Et_2O, 1.3556 and –; EtOH, 1.3633 and 39.10; light ligroine, 1.38 and 39.17; cyclohexane, 1.4264 and –; $CHCl_3$, 1.4486 and 40.72; C_6H_6, 1.5017 and 41.32; pyridine, 1.5085 and –; CS_2, 1.6319 and 44.18.

Abstract 5.—The formula given by the author for $\lambda_{max.}$ of polyenes in the form $(\lambda_{max.})^2 = a + B(1 - C^N)$ is examd. for its susceptibility to the solvent. It is shown that the parameter B is characteristic to each solvent and the following linear relationship to n_D^{20} was found: $B_{calcd.}^{solv.} = 12.976 + 19.019\, n_D^{20}$. The agreement is very good between the observed value and B value calcd. from the above formula for MeOH, hexane, petr. ether, Et_2O, ligroine, cyclohexane, $CHCl_3$, C_6H_6, pyridine, and CS_2. The parameters a and C are independent of solvents. The author calcd. $\lambda_{max.}$ for 57 polyene derivs. in the above solvents and compared them to the values given in the literature.

[1] The abstract that appears just before this in the abstract journal is that for Part I of the original paper and gives the calculation formula for the wave length of absorption maxima.

2. ANALYSIS OF THE ABSTRACTS

Of the foregoing five abstracts, the original of abstract 1 is the author's own abstract in Japanese, printed with the original article. It is therefore an indicative abstract written for the convenience of the reader in judging whether or not to read the whole paper.

Abstracts 2 to 5 were printed in various asbtract journals and are considered to be informative abstracts which present the contents of the original article in considerable detail and give sufficient data in case the original article is not easily available or the language used is not commonly understood.

Besides these two kinds of abstracts, there is another kind which may be termed a "telegraphic-style" abstract. Both the indicative and informative abstracts can be reduced to this telegraphic style, which is often used for storage of information but may infrequently be adapted to formal publication, such as in abstract journals.

Information contained in the original article may be summarized as follows:

(1) (1.1) Absorption maximum of polyenes was found to be represented by the formula $(\lambda_{max.})^2 = a + B(1 - C^N)$
  (1.2) $= A - BC^N$

(2) (2.1) Of the parameters in (1.1), only B changes with the solvent.
  (2.2) General formula is $(\lambda_{max.})^2 = [-2.12 + B(1 - 0.920^N)] \times 10^4\ m\mu^2$.

(3) (3.1) $B^{solv.}$ was determined for seven solvents.
  (3.2) Values of $\lambda_{max.}$ calculated by using the parameter $B^{solv.}$ agreed with observed values published in the literature.
  (3.3) Comparison of values given in (3.2) was made in 57 compounds.

(4) (4.1) Linear relationship exists between $B^{solv.}$ and n_D^{20}.
  (4.2) This linear relationship is represented by $B_{calcd.}^{solv.} = 12.976 + 19.019\, n_D^{20}$.
  (4.3) $B^{solv.}$ in (3.2) and $B_{calcd.}^{solv.}$ in (4.2) are in good agreement.
  (4.4) This agreement was examined in eight solvents.

(5) (5.1) $\lambda_{max.}$ was calculated from information (2.2) by using $B_{calcd.}^{solv.}$ calculated from (4.2), and the values agreed well with observed values published in literature.
  (5.2) This calculation was made on three solvents.
  (5.3) Eight examples are given on six compounds.

Of these "Facts," (2.2) is the formula obtained by substitution of values into the parameters in (1.1) or (1.2) and is synonymous with (1.2) + (2.1) or (1.1) + (2.1). Indeed, none of the abstracts took up (2.2).

Fact (4.2) is the numerical expression of (4.1), and either will be sufficient in the abstract. All the abstracts did mention (4.1).

Table I shows how the five abstracts treated these facts.

The ratio of facts (information) given in Table I disregards the importance of or the number of concepts contained in such facts, and the difference in the length required for expression of a fact, but it is interesting to note that the amount of information contained in a certain number of words tends to decrease with increasing length of an abstract.

Table I
Analysis of the Abstracts

	Indicative abstract	Informative abstract			
Abstract no.	1	2	3	4	5
Fact no.					
(1.1)	–	–	–[a]	–	–
(1.2)	+	+ (error)	–[a]	+	–
(2.1)	+ (error)[b]	+ (error)[b]	+[c]	+ (error)[b]	+
(3.1)	–	+[d,e]	±[e,f,g]	±[d,g]	±[d,e]
(3.2)	+	+	+ (error)[h]	+	+
(3.3)	–	+	+	+	+
(4.1) = (4.2)	+	+	+ (error)	+[i]	+
(4.3)	–	–	–	–	+
(4.4)	–	–	–	–	–
(5.1)	–	–	+	–	+
(5.2)	–	–	±[d,g]	±[d,g]	±[d,g]
(5.3)	–	–	–	–	–
No. of facts[j]	4	6	6	6	8
Relative length of abstract[k]	1.00	1.56	1.63	2.52 (1.73)	2.77
Ratio of facts[l]	4.0	3.8	3.7	2.4 (3.5)	2.9

[a] This is not given in the abstract, but the abstract of the preceding paper (given immediately above this abstract) contains it, so it can therefore be understood. [b] This error was made by the error in the author's abstract. (1.2) is cited instead of (1.1). [c] This abstract is insufficient to judge whether the foregoing error has been corrected or not. [d] Names of the solvents are given. [e] Solvents given in the preceding paper (Part I) are also included. [f] Names of the solvents are given, but one is missing. [g] Informations (3.1) and (5.2) are combined, and no distinction can be made between them. [h] Calculated values is confused with observed values given in published literature. [i] Gives values for all the solvents. [j] Total of plus signs; ± was counted as 0.5. The abstracting errors can be corrected by changing the sentences slightly; this scarcely affects the length of the abstracts. Each abstract was therefore calculated as having been corrected. [k] Ratio of the length of abstracts with that of abstract 1 as 1.00. [l] Ratio of the number of facts contained in an abstract to the length of that abstract. This gives the ratio of the amount of information in a definite length of an abstract.

The exceptionally small ratio of facts in abstract 4 is due to the fact that it contains the value of $B^{\text{solv.}}$ for each solvent; the figures given in parentheses refer to the length of the abstract and the ratio of facts obtained by excluding the values of $B^{\text{solv.}}$ (see Fig. 3).

When the length of an abstract becomes the same as that of the original paper, all the information contained in the latter will be found in the text; this would mean that the number of facts will be 14, length of the abstract 11.6, and ratio of facts, 1.2.

Abstract 5 was prepared from the original article, without reference to the author's abstract, and is accurate.

Abstracts 1 and 5, which are rather opposite in character, may be reduced to the telegraphic style shown below.

Abstract 1 (T).—Absorption max. of substituted polyenes $(\lambda_{\text{max.}})^2 = a + B(1 - C^N)$. B varies with each solvent. B and n_D of each solvent are linear.

Abstract 5 (T).—Author's formula for $\lambda_{\text{max.}}$ of polyenes $(\lambda_{\text{max.}})^2 = a + B(1 - C^N)$. Solvent affects only B. Good agreement between $B^{\text{solv.}}_{\text{calcd.}} = 12.976 + 19.019\ n_D^{20}$ and B from observed values (MeOH, hexane, petr. ether, Et_2O, EtOH, ligroine, cyclohexane, $CHCl_3$, C_6H_6, pyridine, CS_2). Calcd. values of $\lambda_{\text{max.}}$ of 57 compds. in the foregoing solvents were compared with observed values given in published literature.

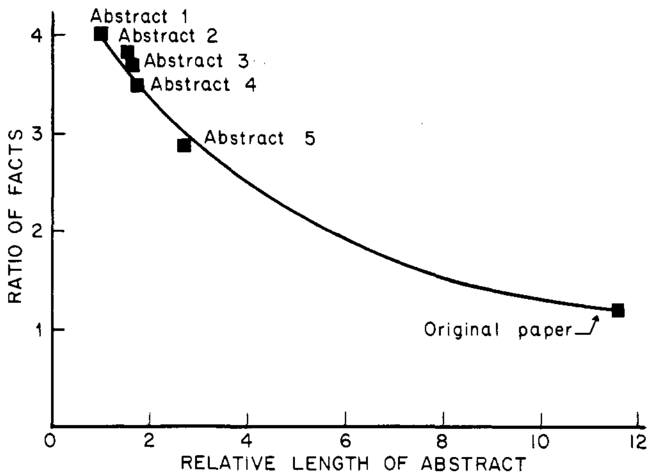

Fig. 3.—Relative length of abstract *vs.* ratio of facts.

These rewritten abstracts, tabulated as in Table I, give the following information, showing that the ratio of facts has been nearly doubled.

	Rewritten abstract 1	Rewritten abstract 5
No. of facts	4	8
Length of abstract	0.55	1.63
Ratio of facts	7.3	4.9

An abstract in telegraphic style containing all the informations except (1.2), (2.2), and (4.2), which are considered to be superfluous for an abstract, will be given below as an example. In this case, the names of the solvents are unnecessary and are not significant even if given.

Example of Telegraphic-Style Abstract with a Large Amount of Information.—U. V. absorption max. of substituted polyenes $(\lambda_{\text{max.}})^2 = a + B(1 - C^N)$. B alone changes with each solvent. Calcd. values of $\lambda_{\text{max.}}$ agree well with observed values in the literature (57 compds. in 7 solvents). Linear relationship between B and n_D^{20}. Values calcd. from n_D^{20} agree well with B from above formula (8 solvents). $\lambda_{\text{max.}}$ calcd. from B obtained from n_D^{20} of 3 solvents agreed well with observed values (8 examples).

This abstract gives the following amount of information: the efficiency is outstanding.

No. of facts	11
Length of abstract	1.76
Ratio of facts	6.2

It may be concluded from this analysis that the amount of information increases with decreasing length of an abstract, and the efficiency increases. A telegraphic-style abstract is easier to prepare and contains a larger amount of information for a given length of an abstract.

ACKNOWLEDGMENT

Grateful acknowledgment is made to Dorothy Mizoguchi for her cooperation in preparing this manuscript.