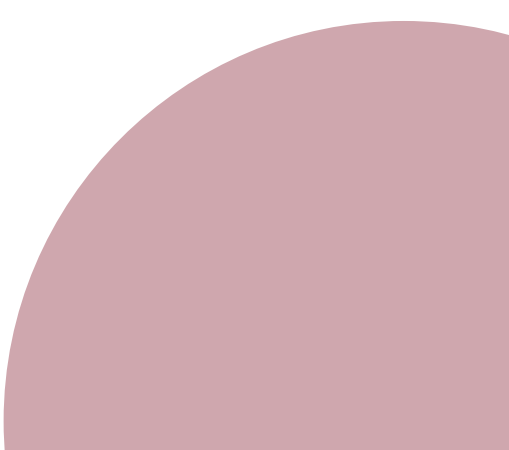


사용자 친화적 타이포그래피 연구

정보 취약계층과 사용자 편의를 위한 최적 디자인

2025-12-20

지도교사 유 미 선생님
중앙고등학교
20731 황태준
범교과 융합 심화탐구



목차

I. 정보 밀도	3
I.1. 언어학적 정보 밀도	3
I.2. 정보 엔트로피에서의 정보 밀도	5
I.3. 시각적 정보 밀도	7
I.4. 통합 정보 밀도	9
II. 정보 밀도 감소를 통한 개선 방안	10
II.1. 글꼴의 가독성	11
II.2. 글자 크기 키우기	12
II.3. 여백 조정	12
II.4. 정보 밀도를 낮추는 어휘와 문체 사용	16
III. 적용 및 실험	17
참고문헌 및 출처	18

초록 컴퓨터 기술은 서양에서 발전해 서양의 언어에 맞게 개발되어 왔다. 지금은 여러 방법으로 다양한 문화권의 언어를 지원하고 있지만, 여전히 기본값은 인도유럽어에 맞추어져 있다. 여기서 제기하는 하나의 문제점은, 특히 한국어에서 영어와의 통사 구조의 차이와 문자의 표의성에서 기인하는 정보 인식의 비효율성과, 낮은 가독성과 시인성, 좁은 공간에 글을 구겨넣기 위해 사용하는 한자어 등으로 인한 낮은 접근성 등의 문제이다. 본 탐구에서는 언어학적 논증과 수치적 계산으로 최적의 文体, 字形, 글자 크기, 字間, 行間 등 수치를 도출하여 이러한 문제에 대한 해결책을 제시하고자 한다.

I. 정보 밀도

정보 밀도란 추상적으로, 어떤 단위 표현 길이 안에 함축된 정보의 양을 말한다. 궁극적으로는 뇌가 일정한 단위 길이나 시간동안 처리해야 하는 정보의 양을 말한다.

I.1. 언어학적 정보 밀도

프랑스 리옹 대학의 연구에 따르면, 각 언어마다 음절 당 정보량(information density per syllable), 즉 ‘정보 밀도’가 다르다[1], [2]. 일반적으로 문자의 표의성이 강할수록 한 문자에 많은 의미가 포함되기에 정보 밀도가 높아진다. 또, 문자의 획수가 많거나 여러 음운을 함축할 경우 당연히 정보 밀도가 높아진다.

예를 들어, 영어 등 인도유럽어족에 속하는 언어는 한 단어를 이루는 문자 수가 많고 표음성이 강해 정보 밀도가 낮다. 말을 빨리 해도 정보 전달 속도는 속도에 거의 무관하게 일정하다. 아래는 영어, 독일어, 러시아어의 예이다*.

All human beings are born free and equal and should be treated the same way.

Alle Menschen sind frei und gleich an Würde und Rechten geboren.

Все люди рождаются свободными и равными в своем достоинстве и правах.

반면, 중국어는 가장 적게는 한 글자, 즉 한 음절이 한 단어이다. 성조가 있어서 소리는 짧으나 담을 수 있는 의미가 네 배가 된다. 한문이 바로 고대 중국어인 점을 생각해보면 쉽다. 그렇기에 중국어는 정보 밀도가 매우 높은 언어에 속한다.

(簡) 人人生而自由平等，应被同等对待。

(繁) 人人生而自由平等，應被同等對待。

*앞으로부터 잠시동안 예시로 드는 문장은 UN의 세계인권선언이다. [3]

*요즘은 일본어를 음절보다 더 세분화된 단위인 모라(mora)가 기본 박자 단위가 되는 모라 박자 언어라고 하는 편이다.

한국어와 일본어는 음절 박자 언어*이며, 중국어의 영향을 받아 한자어가 많고 문자가 표의성을 띄지만 †, 교착어의 특성으로 어간이 어미로서 활용하고 어근이 접사로서 굴절하여 호흡을 조절하는 특성이 있다. 그래서 정보 밀도가 중국어와 인도유럽어 사이에 있으나, 여전히 높은 편이다. 특히 한국어는 일본어와 달리 초성, 중성, 종성을 모두 한 글자에 포함시키고, 표현 가능한 발음이 일본어보다 많아 일본어 보다는 정보 밀도가 높다. 일본어는 대체로 폐쇄 음절 언어로 종성이 없다시피하고‡ 표현 가능한 발음이 적어 청취 시 정보 밀도는 낮은 편이지만, 표기 시 한자의 사용 때문에 쓰여 있는 걸 읽을 때는 정보 밀도가 높다고 할 수 있다. 한국어는 한자로 쓰나 한글로 쓰나 어차피 한 글자이기 때문에 밀도 면에서는 차이가 별로 없다.

모든 인간은 태어날 때부터 자유로우며 그 존엄과 권리에 있어 동등하다.

모든 人間은 태어날 때부터 自由로우며 그 尊嚴과 權利에 있어 同等하다.

すべての人間は、^{にんげん}生れながらにして^{じゆう}自由であり、かつ、^{そんげん}尊嚴と^{けんり}權利とについて^{びようどう}平等である。

§II

여기서 말하는 정보 밀도란 발화를 청취할 때 초당 들어오는 음절 당 정보의 양을 말하는 것이다. 여기서 본 일본어 등 일부 예외를 제외하면 들을 때와 읽을 때의 정보 밀도는 대체로 비례하고, 전체적으로 양의 상관관계를 보인다. 독해 시 정보 밀도는 간략화하여 아래와 같이 구할 수 있을 것 같다.

$$ID \sim \frac{1}{L} \quad (1.1)$$

처음에 단도직입적으로 개념을 도입했으나, 이쯤에서 정보 밀도가 중요한 이유를 명시하고 넘어가겠다.

정보 밀도는 사용자 편의 측면에서 중요하다. 특히 웹페이지나 애플리케이션에 나타난 글을 읽을 때는 이해하는 속도보다 눈알이 줄글을 훑는 속도가 빠를 수 있기에 정확한 이해가 어려워 정보 인식 효율이 떨어지거나 중요 약관의 경우 제대로 숙지하지 못하는 문제가 발생할 수 있다**. 무엇보다, 정보가 가득 들어찬 느낌을 주면, 그 양에 압도당한다거나, 시력이 안 좋은 노인 등의 경우 해독에 어려움이 있는 등 여러 이유로 읽기가 싫고 어려워진다.

*한글은 형태음소적 표음문자인 한글을 사용하고, 일본은 표음문자인 가나(仮名・かな)와 함께 가독성을 위해 한자를 사용한다.

†고대의 일본어에는 정말로 초성과 중성밖에 없었던 것으로 보인다고 한다. 한국이나 중국, 한자문화권의 영향을 받아 중성 모라에 해당한다고 볼 수 있는 촉음(促音), 발음(撥音)이 생겼을 가능성이 높다.

§위에서 보이는「すべて」처럼,「全て」로 한자 표기를 해도 되지만 가독성에 영향이 크게 없는 경우 하지 않기도 한다.

¶한국어와 일본어의 한자 어휘는 비슷한데 한자 자형이 다른 것을 관찰하는 것도 재미있다.

‖한국어 문장은 일본어 문장을 번역해온 티가 나는 것 같다. 사실인지는 모르겠다. 다만, 직역체는 다음과 같아야 할 것이다: 모든 인간은, 태어날 때부터 자유로우며, 또한, 존엄과 권리에 있어 평등하다.

**일반적으로 읽는 것을 의식적으로 멈추지 않는 한 안구가 글줄을 읽으며 회전하는 속도는 이해 여부나 정보 밀도 등에 거의 무관하다. 즉 읽을 때는 줄글의 길이가 시간과 같은 요인으로 작용한다고 할 수 있다.

I.2. 정보 엔트로피에서의 정보 밀도

위에서 알아본 단순한 정보 밀도도 있는 반면, 다른 개념의 정보 밀도도 인지에 영향을 미친다. 새넨 (Shannon) 엔트로피란 이론 컴퓨터과학에서 정보 엔트로피*[4], [5], [6],의 일종으로, 우리의 맥락에서는 다음에 올 의미†가 얼마나 예측하기 힘든지를 의미한다. 즉, 정보의 예측 불가능성이 정보의 엔트로피, 즉 무질서도라는 것이다. 아래 두 문장을 비교해 보자.

나는 배가 고프면 서둘러 허기를 달래기 위해 밥을 먹는다.

나는 고에너지의 몸소 경험해 보고 싶어서 우라늄을 먹는다.

문장의 길이는 비슷하나, 예측하기 어려운 문장은 아래 것이다. 즉 아래 것을 이해하는 데 뇌의 어떠한 처리 자원이 더 많이 소비되기에 정보 밀도가 더 높다는 것이다. 새넨 엔트로피의 식은 아래와 같다.

$$H(X) = - \sum_{x \in X} P(x) \log_2 P(x) \quad (1.2)$$

즉 다음 의미가 나올 확률에 대해 가짓수 N 이 증가하면 엔트로피는 $\log_2 N$ 비율로 증가한다. 타이포그래피에서 다음 단어나 문장의 엔트로피에 따라 뭔가를 조정하기에는 미학적으로 좋지 않고 가독성이 떨어지며 애초에 구현이 비현실적이다. 다만, 한자문화권의 글자들이 라틴 글자들보다 한 글자 단위의 평균적 엔트로피가 더 높다는 것은 짚고 넘어가야 한다.

현대 한국어의 글자 수는 유니코드 완성형 한글 기준 11,172 자이며, 그 중 자주 사용된다고 알려진 EUC-KR의 KS X 1001 완성형 한글은 2,350 자이다. 자주 접할 일은 없지만, 중세국어 자모와 표기법을 도입 시 표현 가능한 유니코드 조합형 옛한글은 자그마치 163,875 자이다. 한글의 경우 자음 14개와 모음 10개로 24자로 충분한 것이니 글자에 대한 엔트로피가 낮은 것이 아니냐고 할 수 있으나, 인간의 뇌는 음절을 처리하는 데 특화되어있고 따라서 한국인을 포함한 한국어 화자는 한글을 글자, 즉 음절 단위로 읽는다. 즉 실제로 음절 단위로 접하게 되는 정보의 가짓수에 대해서는 위의 서술을 따르는게 타당하다고 할 수 있다.

일본어의 경우 히라가나(平仮名), 가타카나(片仮名), 탁음(濁音), 반탁음(半濁音), 스테가나(捨て仮名)‡ 등을 포함해 150 자 내외이며, 일본어 표기에 일반적으로 사용하는 한자는 일본에서 지정한 상용한자(常用漢字)로 총 2136 자이다.

중국어에 사용되는 상용한자는 3,500 자 내외, 실제로 단어 구성 등에 통용되는 한자는 7,000 자 내외로 알려져 있으며, 옛날부터 한문이 그랬듯 정확한 수치는 알 수 없다.

*엔트로피란 무질서도(無秩序度)를 의미하는 물리학, 열역학에서의 용어이지만 통계학, 컴퓨터과학 등에서도 비유적인 의미에서 착안한 개념으로 쓰인다.

†문장, 단어, 글자 등 의미 구분을 가능케 하는 객체를 말하는 것이다.

‡촉음(促音), 요음(拗音), 장음(長音) 등을 표기할 때 사용하는 작은 가나.

즉 많아봤자 30여 개 내외의 글자로 이루어진 인도유럽어와 달리 한중일(CJK) 문자 체계는 한자문화권의 연상적(聯想的) 사고방식과 전각 표의주의, 모아쓰기 표기 때문에 글자 수의 가짓수가 정말 많다. 그러므로 글자 단위의 새넨 엔트로피가 높다고 할 수 있다.

구체적으로, 언어에 대해 그 글자수와 글자수에 대한 새넨 엔트로피의 관계에 대한 식을 생각하자. 엔트로피가 최대가 되는 경우는 가능한 글자 수가 N 개일 때이다.

$$H_{\max}(N) = - \sum_{i=1}^N \frac{1}{N} \log_2 \frac{1}{N} = \log_2 N \quad (1.3)$$

다만, 언어 생활은 실제로는 각 글자가 균등하게 등장하지 않고, 파레토 법칙을 따른다[7]. 파레토 법칙 또는 80/20 법칙이란, 많은 경우에서 전체 결과의 80%가 전체 원인의 20%에서 일어나는 현상을 가리킨다. 언어에서는 전체 의미의 20%가 전체 빈도의 80%를 차지하는 경향을 보이는 것으로 발현한다. 즉 뇌는 자주 쓰는 글자에 이미 익숙해져 있고 그 빈도가 높으므로 실제로는 H_{\max} 값보다 정보 엔트로피가 낮을 것으로 예상할 수 있다. 즉 자주 쓰는 글자가 상위 20%라고 가정하면 실제로 뇌가 체감하는 유효 탐색 공간은 글자 수 N 이 아닌 $0.2N$ 이므로 식을 보정해야 한다. 나머지 80% 글자의 등장 빈도 20%는 어림잡아 무시한다고 하면 유효 정보 엔트로피는 아래와 같다.

$$H_{\text{eff}} \approx \log_2(0.2N) = \log_2 N + \log_2 0.2 \approx \log_2 N - 2.3219 \quad (1.4)$$

즉 (당연히) 값이 감소함을 알 수 있다. 경향을 알았으니 더 정확한 계산을 위해 나머지 20%를 고려하여 구하자. 그렇게 하기 위해 전체 빈도의 80%를 차지하는 상위 20%의 글자를 A 군(群), 전체 빈도의 20%를 차지하는 하위 80%의 글자를 B 군이라고 하면, A의 글자 수는 $0.2N$, 각 글자의 등장 확률은 $P_A = 0.8/(0.2N) = 4/N$ 이고 B의 글자 수는 $0.8N$, 확률은 $P_B = 0.2/(0.8N) = 1/4N$ 이 된다.

이때 두 군을 고려한 엔트로피 식은 이렇게 된다.

$$\begin{aligned} H_{\text{eff}} &= 0.2N(-P_A \log_2 P_A) + 0.8N(-P_B \log_2 P_B) \\ &= 0.2N \left(-\frac{4}{N} \log_2 \frac{4}{N} \right) + 0.8N \left(-\frac{1}{4N} \log_2 \frac{1}{4N} \right) \\ &= (0.8 \log_2 N - 1.6) + (0.2 \log_2 N + 0.4) \\ &= \log_2 N - 1.2 \end{aligned} \quad (1.5)$$

$$\boxed{\therefore H_{\text{eff}} = \log_2 N - 1.2} \quad (1.6)$$

갑자기 엔트로피가 1이나 증가하는 것을 볼 수 있다. 이는 자주 쓰지 않는 글자가 등장하게 되면 그만큼 더 예측 불가능성이 증가한다는 뜻이 된다.

어쩌다가 로그의 밑의 2와 80, 20에 0.2, 0.8을 곱하는 것 때문에 로그 값이 예쁘게 떨어져서 나왔지만, 여전히 80/20 법칙이 정확히 80% 20% 비율로 떨어지는 것이 아니라는 걸 감안하여 이 조차도 근사치라는 것을 염두해야 한다. 하지만 이것보다 더 일반적이면서도 합리적인 접근을 하기는 어려울 것 같아 이대로 진행하도록 하겠다.

이것으로 영어, 독일어, 러시아어, 한국어, 일본어, 중국어의 유효 정보 엔트로피 H_{eff} 를 구해보자. 산출 공식은 $\log_2 N - 1.2$ 이며, 소수점 둘째자리에서 반올림했다. 영어는 알파벳 26 자, 독일어는 알파벳 26 자에 움라우트(äöü, Umlaut) 3 자, 에스체트(ß, Eszett/Scharfes S)로 30 자, 러시아어는 키릴 문자 33자, 한국어는 KS X 1001 완성형 한글 2350 자, 일본어는 히라가나 46 자, 가타카나 46 자, 상용한자 2136자로 2228자, 중국어는 상용한자 3500 자와 실용한자 7000자 간 임의의 적정값 5000 자로 기준을 잡았다.

Table 1.1 — 언어별 문자 수(N)에 대한 유효 정보 엔트로피 산출표

언어	문자 수 (N)	유효 엔트로피 (H_{eff})
영어	26	3.5
독일어	30	3.71
러시아어	33	3.84
한국어	2350	10
일본어	2228	9.92
중국어	5000	11.09

위 표 Table 1.1 에서 보이듯, 인도유럽어와 CJK 간에는 말도 안 되는 엔트로피 차이가 있다.

I.3. 시각적 정보 밀도

시각적 정보 밀도란 그냥 만들어낸 말로, 글자에서 단위 면적 당 획이 얼마나 많은가를 의미한다. 사실 시력이 좋지 못한 사람들이나, 작은 화면을 들여다 보는 사람들, 또는 요즘에는 그냥 읽는 것을 싫어하는 사람들에게 중요한 것은 의미보다도 눈에 보이는 복잡도이다.

어떤 전광판에 같은 의미의 글자를 서로 다른 언어로 같은 크기로 띄운다고 생각해 보자. ‘출구’ 라던가 ‘쓰레기 버리지 마시오’ 같은 것은 흔한 표지이므로 통채로 인식하거나, ‘쓰레기’처럼 앞글자만 읽고 넘기는 등 글을 읽는 것이 아니라 하나의 상(想)처럼 보고 넘겨 버리니 좋은 예시로서 적절하지 않고, 좀 독특한 표지판을 가져와 보겠다.

Figure 1.1 — 진짜 있는 표지판이다.



이것을 각국 언어의 적절한 표지판 문체로 번역해 보았다.

Table 1.2 — 동일 의미에 대한 언어별 표기 비교

언어	표지판 번역
영어	CAUTION: FALLING COWS
독일어	ACHTUNG: FALLENDE KÜHE
러시아어	ОСТОРОЖНО: ПАДАЮЩИЕ КОРОВЫ
한국어	주의: 소 낙하
일본어	注意：牛の落下
중국어(間)	注意：高空坠牛
중국어(繁)	注意：高空墜牛

앞서 확인한 정보 밀도의 특성 뿐 아니라, 획이 많을수록 뾰뾰하게 보여서 읽기 싫어진다는 것을 느낄 수 있다. 다만 문자가 뾰뾰해서가 아니라, 키릴이나 가나 등 생소한 문자가 익숙하지 않게 보여서 읽기에 거부감이 드는 것일 수도 있으므로 아래의 예시도 참고한다.

폭 넓고 뾰뾰한 숲을 뚫다	둘이 나란히 서서 웃다
----------------	--------------

두 문장의 글자 수는 비슷하지만 각 글자의 평균 획 수가 달라 시각적 밀도가 다르다는 것을 알 수 있다. 특히, 각 단어를 하나의 덩어리로 외우고 있기에 읽기 쉬운 이런 예시와 달리 각 단어가 생소하다면

더욱 더 어려울 수 있다. 이는 온라인 상황에서는 IT 용어나 신조어, 유행어를 잘 알지 못하며 시력이 좋지 않은 노인들에게 큰 영향을 줄 수 있다.

시각적 밀도가 높은 극단적인 예시로는, 장난이나 농담으로 쓰이기도 하는 이런 글자들을 들 수 있다.

鯢魚癰*

[illegible]

또한, 현대에 일반적으로 통용되는 서식에서 가독성이 심각하게 떨어지는 문자 체계의 예시 중 하나로 옛한글이 있다.

나.랏:말쑥.미

中 國 · 국 · 에 달 · 아 文 字 · 종 · 와 · 로 서 르 스 못 · 디 아 · 니 呼 · 씩

·이런전·츠·로어·린百·빅姓·성·이니르·고·저·홀·배이·셔·도

ㅁ·츄:내제·쁘·들시·러퍼·디:몬훙·노·미하·니·라

·내·이·를·爲·왕·하·야·어·엇·비·너·겨·새·로·스·믈·여·들·字·쥬·를·땡·고·노·니

·사·름·마·다·히·여·수·빙·니·겨·날·로·부·매·便·安·한·키·고·저·흥·썩·르·미·니·라·

一 訓民正音 御製序文

보이듯, 옛한글은 더욱 많은 가능한 글자 수 뿐 아니라, 띄어쓰기의 부재, 글자의 많은 획수와 좌가점(左加點)의 표기, 국한문혼용(國漢文混用) 때문에 여러모로 정보 밀도가 높으며, 서양식 가로쓰기보다는 동양식 세로쓰기에 적합함을 알 수 있다.

시각적 밀도를 표현할 수 있는 식은, N 을 획 수, A 를 면적이라고 하면 다음과 같다고 할 수 있겠는데,

$$\text{VD} = \frac{N}{A} \quad (1.7)$$

이때 이 VD 값이 특정 임계치를 넘으면 읽기 어려워지고 피곤해진다.

I.4. 통합 정보 밀도

이 세 가지 정보 밀도를 합쳐 생각할 수 있도록, 통합 정보 밀도라는 임의로 정한 양을 합리적으로 유도하여 정의하겠다. 문자는 편한 대로 임의로 다시 놓겠다.

먼저 언어학적 정보 밀도는 $L = 1/l$ 에 비례했다. 또, 유효 정보 엔트로피는 가능한 전체 글자 수 N 에 대해 $H = \log_2 N - 1.2$ 로 하기로 했다. 또 시각적 정보 밀도는 글자 당 획 수 s 와 글자 당 차지하는 면적 a 에 대해 $S = s/a$ 였다. 이때 이 정보 부하량을 함께 생각한 통합 정보 부하량(total cognitive load) Λ 는 아래와 같다고 할 수 있다.

*마지막 한자 𪎭는 국수이름/방방면 방 이라는 한자로 총 57획이며, 강희자전(康熙字典)에도 등장하는 한자이다.

$$\Lambda = HSL \quad (1.8)$$

이것을 전체 글이 차지하는 시각적 면적 $A = wh$ 로 나누면 통합 정보 밀도가 나올 것이다.

$$\Psi = \frac{\Lambda}{A} = \frac{HSL}{A} = \frac{1}{A} \frac{s(\log_2 N - 1.2)}{al} \quad (1.9)$$

베버-페히너(Weber-Fechner) 법칙[8]에 따라 인간이 느끼는 감각은 자극의 로그에 비례한다. 따라서, 실제로 느끼는 인지 부하 P 는 정보 밀도 Ψ 의 로그에 비례할 것이다.

$$P \propto \ln \Psi \quad (1.10)$$

결론적으로,

$$P \propto \ln \left(\frac{1}{A} \frac{s(\log_2 N - 1.2)}{al} \right) \quad (1.11)$$

여기서

- A : 글 전체가 차지하는 시각적 면적
- a : 글자 하나 당 차지하는 면적
- l : 단어 하나 당 평균 길이
- s : 글자 당 획 수
- N : 가능한 모든 문자의 가짓수

II. 정보 밀도 감소를 통한 개선 방안

이렇게, 언어학적 정보 밀도, 정보 엔트로피, 시각적 정보 밀도를 모두 고려하였을 때, 특히 웹 환경에서 정보 밀도가 뇌가 편안하게 처리할 수 있는 문턱 정보 밀도보다 높으면 생길 수 있는 문제는 다음과 같다.

- 줄글을 읽었지만 무슨 소리를 해 놓은 것인지 잘 모르겠다.
- 뻑뻑한 글자를 홑뜨고 보느라 눈이 아프다.
- 애초에 줄글을 보면 읽기가 싫다.
- 시력이 약한 경우 더욱 읽기가 싫을 뿐 아니라 읽을 수 없다.

이러한 문제를 해결하고 특히 정보 취약 계층을 위해 온라인 상의 CJK 문자 가독성, 시인성, 접근성 향상을 위해 고려해볼 수 있는 해결책은 아래와 같다.

- 가독성 좋은 글꼴 선택
- 자간(자간), 행간 여백 조정
- 글자 크기 키우기

- 정보 밀도를 낮출 수 있는 어휘 사용
- UI/UX 요소의 적절한 배치

당연한 것들이라고 할 수 있으나, 문제는 이러한 조치는 대부분 웹페이지의 미학적인 가치를 떨어뜨린다. 수학적으로 다루어 볼 수 있는 것들 위주로 자세히 알아보겠다.

II.1. 글꼴의 가독성

글꼴은 그 겉모양을 기준으로 했을 때 크게 명조/부리(serif), 고딕/민부리(sans serif), 손글씨(caligraphy)의 세 가지로 나눌 수 있다.

서양에서 구텐베르크 Fraktur 활자인쇄체가 Roman 명조체에 표준 인쇄체 자리를 내준 이후, 명조는 신뢰성, 정통성, 단단함 등을 상징하게 되었다. 또, 공식적인 문서에서는 활자 인쇄의 전통을 따라 명조체를 사용한다. 명조체의 특징은 획에 부리(serif)가 달려 있다는 것인데, 이것이 명조를 멋있는 글꼴로 만들면서도 가독성을 해치는 주 원인이다.

고딕체는 미니멀리즘(minimalism)을 추구하는 모던 디자인에 부합하는 컴팩트한 글꼴로, 부리가 없다. 지루해 보인다는 단점이 있지만, 특징이 없는만큼 변형을 가하기 쉽고 가독성이 뛰어나다는 장점때문에 특히 웹 디자인에서 즐겨 사용된다.

일반적으로 가독성이 좋은 순서대로 글꼴을 나열하면 고딕 - 명조 - 손글씨 이다. 직접 비교해보자*.

다람쥐 쳃바퀴 quick brown fox 道吾善者是吾賊 言い訳はいらない
 다람쥐 쳃바퀴 quick brown fox 道吾善者是吾賊 言い訳はいらない
 다람쥐 쳃바퀴 quick brown fox 道吾善者是吾賊 言い訳はいらない

가독성을 위해서라면 무조건 고딕을 써야 하겠지만, 가독성보다 우선시되는 브랜드 이미지나 상품성, 또는 단순한 디자인의 아름다움 등 때문에 그럴 수 없다. 손글씨는 어차피 특수한 곳에서만 쓰이며, 가독성을 별로 중요시하지 않기에 고려하지 않더라도, 명조체는 그 이미지를 양보할 수 없기에, 가독성을 조금이라도 향상하는 방식으로 이미 진화해 왔다.

다람쥐 헌 쳃바퀴에 타고파.
 다람쥐 헌 쳃바퀴에 타고파.
 다람쥐 헌 쳃바퀴에 타고파.
 다람쥐 헌 쳃바퀴에 타고파.

*글꼴은 위에서부터 본고딕, 본명조, LXGW WenKai.

다람쥐 헌 쳇바퀴에 타고파.

다람쥐 헌 쳇바퀴에 타고파.

위에서부터 바탕, HY신명조, KoPub 바탕, 네이버 나눔명조, 본명조, 조선일보명조이다. 예전부터 사용하던 한국어 명조인 바탕과 신명조에서부터, 획의 굵기 차이를 줄이거나 불필요한 부리를 제거하거나 자음을 더 돋보이게 만들고 곡률을 줄이는 등의 방법으로, 명조의 느낌을 살리면서도 가독성을 향상하는 시도를 한 것을 확인할 수 있다. 글꼴 디자인은 아주 많은 노력을 요하기에, 가장 쉽게 할 수 있는 것은, 명조 글꼴을 사용하려고 할 때는 목적에 맞는 형태를 가진 글꼴을 고심하여 고르는 수밖에 없겠다.

II.2. 글자 크기 키우기

어디까지나 키울 수 있다면 좋겠지만 그렇게 한다면 미학적인 것을 아예 망치게 될 가능성이 크다. 또, CJK 글꼴과 라틴 글꼴을 옆에 놓고 비교해 보면, 별도의 조정을 하지 않는 경우 CJK 글꼴과 라틴 글꼴보다 같은 글꼴 크기를 지정했음에도 더 큰 것을 확인할 수 있다. 본 보고서에서는 의도적으로 CJK 폰트의 크기를 줄여 라틴 폰트와 크기를 맞추었다. 이것은 크기보다 더 중요한 요인이 있다고 생각한 결과이기도 하고, 웹 디자인이 아닌 지면 문서이기 때문에 이 상황에서 가독성의 대부분은 글꼴과 여백에서 온다고 판단했기 때문이기도 하다.

II.3. 여백 조정

行間은 익히 알고 있으니 따로 설명하지 않는다.

字間은 가장 등한시되는 타이포그래피의 요소이다. 정보 밀도의 측면이 아니더라도, 서양에서도 특히 로고 글자의 자간을 조금만 조정하는 것만으로 엄청난 디자인 효과를 본 사례가 존재한다.

인도유럽어는 정보 밀도가 낮기 때문에, 특수한 상황이 아닐 때 자간이 넓으면 읽는 사람에게 마치 말을 굼벵이처럼 하는 느낌을 주게 된다. 일반적인 자간으로도 부족한 그들은 아예 그들의 글꼴 디자인에 자간을 찌부시키기 위한 맞춤형 기능들을 넣어두었다. 아래는 한국어 명조 글꼴로 쓴 라틴 문자와 서양 명조 글꼴로 쓴 라틴 문자의 비교이다. 위의 것이 HY신명조, 아래 것이 Times 계열 글꼴인 STIX Two이다.

The quick brown fox jumps over the lazy dog.

The quick brown fox jumps over the lazy dog.

먼저, 기본적으로 각 알파벳의 크기는 비슷함에도 STIX의 자간이 좁아 신명조로 쓴 문장의 길이가 더 긴 것을 알 수 있다. 또, 'fox'의 'f'와 'o' 사이를 유심히 보면, f의 구부러지는 꼭지 부분이 o의 상단 공간을 침범한 것을 알 수 있고, 'jumps'의 j가 그에 선행하는 여백에 침범하는 것을 볼 수 있다. 그

외에도, 잘 만들어진 라틴 글꼴라면 갖추어야 할 여러 기능들이 자간을 최대한 줄이는 데 초점이 맞추어져 있다. 어디에서 그러한 기능이 적용되었는지 찾아보자.

scoff firm Reykjavíkurborg Jaworzyna Śląska

같은 글꼴 내에서도, 그리스 문자나 키릴 문자 등 정보 밀도가 높은 것들은 이렇게 찌부시키는 기능이 없는 것을 알 수 있다.

Вишгород Вашківці Κόρινθος

이렇게, 실존하는 예시를 보았을 때 정보 밀도 이론은 타당하게 받아들여질 수 있다는 것을 알 수 있다.

다만, 동아시아 CJK 문자들은 아까 보았듯, 일반적인 라틴 문자의 3 배가 넘는 정보 밀도를 가지고 있다. 물론 우리 문자의 글꼴에도 정보 밀도를 분산하기 위한 노력이 없는 것은 아니다. 먼저, 동아시아 문자에는 전각(全角, full-width)과 반각(半角, half-width)이라는 개념이 있으며, 한글 완성자와 한자, 히라가나는 언제나 전각이다. 상대적으로 직선적이라 봐줄만한 가타카나만이 반각으로 쓰일 수 있다. 즉, 우리 문자들은 기본적으로 라틴 문자보다 한 글자당 넓이가 넓다. 예시를 보자. 처음부터 한글, 한자, 히라가나, 전각 가타카나, 반각 가타카나이다. 그 옆에는 비교를 위한 반각, 전각 라틴 문자도 두었다*.

한글 漢字 ひらがな カタカナ 𐆑𐆒 Half Full

이렇게 라틴문자보다 글자 당 면적이 넓어 어느 정도 높은 정보 밀도를 상쇄할 수 있다. 하지만 과연 이대로 충분한가? 답부터 하자면 아니다. 앞서 보았던 밀도의 개념을 통해 이것을 증명해 보이고, 그렇다면 자간의 최적값은 무엇인지 도출하겠다.

라틴 문자의 대표로 영어를 선택하고, 일단 한국어에 대한 결론을 도출해 보기 위해 한국어를 그 비교 대상으로 삼겠다. 현재 타이포그래피 환경에서 라틴 문자가 정보 밀도를 적절하게 조정하여 최적의 가독성을 가지고 있다고 가정할 때, 한국어는 얼마만큼의 여백을 가져야 라틴 문자와 같은 정보 부하를 독자에게 줄 수 있는가를 구해 보겠다.

정보 엔트로피는 앞서 구했 놓았으니, 정보 엔트로피를 전개하지 않고 통합 정보 부하량의 계산식을 제시하면 이렇다.

$$\Lambda = \frac{Hs}{al} \quad (2.1)$$

우리가 늘려야 할 여백 계수 R 를 구하기 위해 써야 할 계산식은

*동아시아 IME가 나온 이후, 전각 문자와 정렬을 맞추기 위한 전각 영문자도 생겼다.

$$R = \ln \frac{\Lambda_{\text{target}}}{\Lambda_{\text{base}}} \quad (2.2)$$

우리의 상황에서는 base가 영어이고 target이 한국어이다.

먼저 앞서 구해왔던 기준 값들을 선언하겠다. 영어의 경우 유효 정보 엔트로피 $H_{\text{en}} = 3.5$ 였다. 평균 획 수의 경우*, 1 획인 글자는 C, J, O, S, U, V, W, X, Z†, 2 획인 글자는 B, D, G, K, L, P, Q, R, T, Y, 3획은 A, E, F, H, I, M, N, W, 4획은 E 이고 대소문자 여부에 따라 조금 차이가 있다. 이렇게 해서 평균을 구해보면, 평균 획 수는 2에서 3 사이로 나온다. 따라서 평균 획 수 $s_{\text{en}} = 2.5$ 로 한다. 단어 당 평균 글자 수는, 신뢰할만한 출처는 찾을 수 없으나 여러 곳을 조사한 결과 약 4.7 자에서 6 자 사이로 한다고 한다. 기준의 차이가 있는 듯하다. 여기서는 $l_{\text{en}} = 5$ 로 하겠다. 따라서 Λ_{en} 은

$$\Lambda_{\text{en}} = \frac{3.5 \times 2.5}{5} = 1.75 \quad (2.3)$$

한국어의 경우 $H_{\text{ko}} = 10$ 이다. 획수 s_{ko} 는 한글 글자 수가 너무 많아 구하기 어려우니, 현대 한글에서 가능한 획수의 최솟값인 2와 최댓값인 20의 평균 11로 하겠다. 완전히 믿을만한 출처는 아니지만, 산업번역명 관리자에 따르면 평균적으로 한국어의 단어 당 평균 글자 수는 3 에서 4 자 정도라고 한다 [9]. $l_{\text{ko}} = 3$ 으로 한다. 이때 전각이 반각의 두 배를 차지한다는 점을 이용해 $a = 2$ 로 두면 Λ_{ko} 는

$$\Lambda_{\text{ko}} = \frac{10 \times 11}{2 \times 3} \approx 18.34 \quad (2.4)$$

이제 늘려야할 면적 배수를 구할 수 있다.

$$R_{\text{ko}} = \ln \frac{18.34}{1.75} \approx \ln 10.48 \approx 2.35 \quad (2.5)$$

즉, 한국어 글은 전각, 반각의 글자 당 면적 차이를 고려하고 보아도 영어보다 글자가 약 2.35 배의 면적을 더 써야 정보 부하의 정도가 비슷하다는 것이다.

보통 가독성이 좋다고 여겨지는 웹 표준에서 영어는 반각(0.5 em)에 행간 1.2를 적용한다‡. 즉,

$$A_{\text{en}} = 0.5 \times 1.5 = 0.75 \quad [\text{em}^2] \quad (2.6)$$

한국어의 시각적 면적은 그러므로 아래와 같이 되어야 한다.

$$A_{\text{kr}} = A_{\text{en}} \times 2.35 \approx 1.7625 \quad [\text{em}^2] \quad (2.7)$$

*이 ‘획’이 무조건 시각적 밀도에 비례하는 것은 아니라는 것을 깨달았지만 획보다 시각적 밀도를 측정하는 더 좋은 기준이 없다고 생각되었다.

†J, W, X는 1획 또는 2획으로 쓰인다. 앞으로 이러한 글자들은 임의로 각 획수 아래 분배한다.

‡타이포그래피에서 em이란 한 글자를 나타내는 단위이다. 화학의 pH와 비슷한 느낌으로, em이 역사적으로 무엇의 약자인지는 불분명하다. 글자의 크기에 따라 상대적으로 서식을 지정할 수 있어 유용하다.

한글 전각은 1 em이므로 이것을 다시 자간 k 와 행간 h 로 풀어 쓰면 자간과 행간은 아래 식을 만족해야 한다. 물론 넣은 값들에 따라 최종 값이 달라지므로 이 식이 완전히 정확한 것이라고 할 수 없지만, 여전히 목표하는 값은 이 근처에서 있어야 할 것이다.

$$A_{kr} = (1 + k)h = 1.7625 \quad [\text{em}^2] \quad (2.8)$$

보통 우리나라 출판업계에서 소설은 행간 1.7 em을 많이 사용한다. 그렇게 할 경우 자간은 0.036 em 정도가 된다. 하지만 출판할 상황이 아닌 경우 1.7은 과한 경향이 있다. 행간을 1.65로 둘 경우 자간은 0.068 정도이고, 한컴오피스 한글 표준인 1.6을 행간으로 할 경우 자간은 0.1 정도가 되어야 한다. 모두 실험해 보자.

(行:1.7em, 字:0.036em) 그의 아내가 기침으로 쿨룩거리기는 벌써 달포가 넘었다. 조밥도 굶기를 먹다시피 하는 형편이니 물론 약 한 첩 써본 일이 없다. 구태여 쓰려면 못쓸 바도 아니로되 그는 병이란 놈에게 약을 주어 보내면 재미를 붙여서 자꾸 온다는 자기의 신조(信條)에 어디까지 충실하였다. 따라서 의사에게 보인 적이 없으니 무슨 병인지는 알 수 없으되 반듯이 누워 가지고 일어나기는 새로 모로도 못 눕는 걸 보면 증증은 증증인 듯. 병이 이대도록 심해지기는 열흘전에 조밥을 먹고 체한 때문이다.

(行:1.65em, 字:0.068em) 그의 아내가 기침으로 쿨룩거리기는 벌써 달포가 넘었다. 조밥도 굶기를 먹다시피 하는 형편이니 물론 약 한 첩 써본 일이 없다. 구태여 쓰려면 못쓸 바도 아니로되 그는 병이란 놈에게 약을 주어 보내면 재미를 붙여서 자꾸 온다는 자기의 신조(信條)에 어디까지 충실하였다. 따라서 의사에게 보인 적이 없으니 무슨 병인지는 알 수 없으되 반듯이 누워 가지고 일어나기는 새로 모로도 못 눕는 걸 보면 증증은 증증인 듯. 병이 이대도록 심해지기는 열흘전에 조밥을 먹고 체한 때문이다.

(行:1.6em, 字:0.1em) 그의 아내가 기침으로 쿨룩거리기는 벌써 달포가 넘었다. 조밥도 굶기를 먹다시피 하는 형편이니 물론 약 한 첩 써본 일이 없다. 구태여 쓰려면 못쓸 바도 아니로되 그는 병이란 놈에게 약을 주어 보내면 재미를 붙여서 자꾸 온다는 자기의 신조(信條)에 어디까지 충실하였다. 따라서 의사에게 보인 적이 없으니 무슨 병인지는 알 수 없으되 반듯이 누워 가지고 일어나기는 새로 모로도 못 눕는 걸 보면 증증은 증증인 듯. 병이 이대도록 심해지기는 열흘전에 조밥을 먹고 체한 때문이다.

— 현진건, 운수 좋은 날 中

한글을 읽기에는 이것들이 시원시원하고 좋은 것 같다. 하지만 구분을 위해 넣어둔 ‘em’ 부분을 보면, 자간이 0.068 만 되어도 간격이 이질감이 들 정도로 멀게 느껴진다. 이것이 CJK 문자와 라틴 문자의 정보 밀도의 차이가 무의식적으로 드러나는 것이라고 생각된다. 실무적으로 구현할 때는 regex 등을 사용해 특정 문자에 특정 서식을 각각 달리 적용하는 방식으로 해야할 것이다.

바로 보듯, 여백은 가독성에 지대한 영향을 미친다. 이런 식으로, 먼저 R 을 구해두고, 업계 표준으로 사용되는 행간을 기준으로 자간을 도출하게 되면 미학적으로 불쾌하지 않으면서도 웹 상에서도 가독성이 좋은 서식을 만들 수 있게 된다!

II.4. 정보 밀도를 낮추는 어휘와 문체 사용

어휘나 문체에 대한 것이 보고서의 주 주제는 아니나, 짚고 넘어가겠다.

먼저, 웹 서비스에서 사용되는 어휘들, 특히 버튼에 사용되는 것들은 점유 공간을 줄이기 위해 짧은 것들로 선택된다. 예를 들어, 확인, 취소, 무시, 동의, 허용 등이 있다. 이런 것을 사용하는 것이 익숙한 사용자에게는 괜찮을 수 있으나, 처음 접하는 사용자에게는 어려울 수 있다. 왜냐하면 애초에 이런 어휘들은 추상적이기 때문인데, VBS 등을 만져보았으면 알겠지만, 옛날 운영체제 수준에서는 선택지가 확인, 확인/취소, 확인/취소/무시 등등 몇 가지 대화상자밖에 띄우지 못했었다. 이런 창들이 여러 기능을 모두 포괄해야 했으므로 어쩔 수 없었던 것이다. 하지만 지금은 상황이 다르다. 그때와 달리 기술이 발전했고, 전자기기는 전문직이나 업계 종사자의 전유물이 아니라 모두의 일상에 스며들어 있다. 그러므로 관습적인 어휘도 바꾸어야 한다. 이미 토스 등 몇몇 서비스는 쉬운 어휘로 바꾸는 것을 추구하고 있다.

예를 들어 아래와 같은 개선 사항이 있을 수 있다.

익일(翌日) - 내일, 금일(今日) - 오늘, 매도(賣渡) - 팔기, 매수(買受) - 사기, 이체(移替)/송금(送金) - (돈) 보내기, 수취인(受取人) - 받는 분/받는 사람, 미가용(未可用) - 사용할 수 없음

이처럼 상대적으로 어려운 한자어를 쉬운 한자어나 순우리말로 바꾸고 풀어 쓰면 어휘의 모호성과 추상성 뿐 아니라 정보 밀도도 낮추는 효과가 있다.

문체에 대해서도 잠깐 서술하겠다. 한국어와 일본어는 SOV(주어 - 목적어 - 서술어) 구조를 가지고 있어, 끝까지 읽어야 의도를 알 수 있다는 약점이 있다. 그래서 속담 중에 한국말은 끝까지 들어야한다는 말도 있지 않은가? 예시를 들어 보면 아래와 같은 문장들이 있겠다.

1. 등록하신 카드가 만료되었사오니 새로운 카드로 교체해 주십시오.
2. 이 버튼을 누르면 데이터가 모두 삭제되오니 주의하십시오.

1 번의 경우, 문장을 끝까지 읽기 전까지 뭘 하라는 것인지 알 수 없다. 2 번의 경우 한국어에서 서술어가 뒤에 나오는 특징과 부정어가 종종 뒤에 나오는 특성 때문에 특정 상황(프로그램 설치를 한다던가...)에서 앞에 이 버튼을 누르- 까지만 보고 버튼을 눌렀다가 낭패를 보는 수가 있다.

꼭 웹 디자인 이야기는 아니지만, 극단적이면서도 좋은 예시가 있어서 가져왔다...

A: 준비됐습니다, 보스. 누굴 쏠까요?

B: 하얀 옷을 입은 남자..

(A가 저격소총을 격발해 하얀 옷을 입은 남자를 사살한다.)

B: 옆에 있는 파란 옷을 죽여라.

— Wall Su, 한국말은 끝까지 들어봐야 하죠 [10]

여기서는 서술어가 뒤에 오고 그에 따라 부사어보다 목적어가 뒤에 오기 때문에 생긴 오해이다. 영어로 했다면 보스가 아래와 같이 말했을테니 이러한 문제가 생기지 않았을 것이다.

B: Kill the blue shirt standing next to the person wearing white clothes.

따라서 이런 식으로 문장을 나눠서 개괄식으로 쓰더라도 중요한 내용을 앞에 나오게 해야 할 것이다.

1. 새로운 카드로 교체: 등록하신 카드가 만료되었습니다. 새로운 카드로 교체해 주세요.
2. 주의! 데이터 삭제: 이 버튼을 누르면 데이터가 모두 삭제됩니다.
3. 파란 옷을 죽여라, 하얀 옷을 입은 남자 옆에 서 있다.

이렇게 하면 글이 길어지고 안 예뻐지기 때문에 널리 받아들이기 어려울 수도 있겠지만 여전히 중요한 부분에는 이러한 기법을 적용해야 할 것이다.

다음과 같은 방법을 제시해 본다.

- 모든 대화상자(모달)의 내용은 동사로 시작한다. “정말로 삭제하시겠습니까?” 대신 “삭제 확인: 정말로 삭제하시겠습니까?” 같이 바꾸고 버튼도 “예”, “확인” 등 대신 “삭제하기”를 명시한다.
- 정보 밀도가 높은 한자어나 획이 복잡한 글자의 비중을 파레토 법칙을 역이용해 20% 이하로 유지한다.

III. 적용 및 실험

도출한 결과를 실감하기 위해 타이포그래피 수치를 테스트하고 언어별 최적 수치를 빠르게 도출하는 웹 서비스를 만들었다. 정적 프론트엔드 웹사이트이며, 프레임워크는 Svelte와 Vite를 사용했다. 기능은 크게 세 가지이다.

1. 타이포그래피 실험실: 자간, 행간, 굵기, 크기 등을 언어별로 조정해보고, 최적 수치를 적용했을 때 가독성을 실감할 수 있다.

2. 통합 정보 밀도 계산기: 앞서 보았던 매개변수 H, a, l, s, w 를 입력하면 결과로 h, k 와 Λ, R, A 를 내놓는 계산기이다.
3. UX 체험: 여기서 제시한 개선사항을 적용한 UI와 그렇지 않은 UI를 비교하여 그 효과를 실감하게 하는 체험 기능이다.*†

Github 리포지토리는 <https://github.com/tjwhang/OptimalCJK>이며, <https://tjwhang.github.io/OptimalCJK/> 에서 웹사이트를 이용해볼 수 있다.

참고문헌 및 출처

- [1] Catherine Matatic, “Human speech may have a universal transmission rate, 39 bits per second.” [Online]. Available: <https://www.science.org/content/article/human-speech-may-have-universal-transmission-rate-39-bits-second>
- [2] Christophe Coup , Yoon Mi Oh, Dan Dediu, and Fran ois Pellegrino, “Different languages, similar encoding efficiency, Comparable information rates across the human communicative niche,” Sept. 04, 2019. [Online]. Available: <https://www.science.org/doi/10.1126/sciadv.aaw2594>
- [3] “Universal Declaration of Human Rights.” [Online]. Available: <https://www.ohchr.org/en/human-rights/universal-declaration/>
- [4] 3Blue1Brown(Grant Sanderson), *Solving Wordle using information theory*. [Online Video]. Available: <https://youtu.be/v68zYyaEmEA?si=V4Lk7VC1T6t9-P1R>
- [5] D. J. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2014.
- [6] C. E. Shannon, “A Mathematical Theory of Communication,” Oct. 1948.
- [7] Juran, “Pareto Principle (80/20 Rule) & Pareto Analysis Guide.” [Online]. Available: <https://www.juran.com/blog/a-guide-to-the-pareto-principle-80-20-rule-pareto-analysis/>
- [8] *Encyclop dia Britannica*, 11th ed., vol. 28. Cambridge University Press, 1911, pp. 458–449.

*처음에는 공평하게 만들고 타이포그래피만 차이를 주려고 했으나, 개선안이 효과가 좋을 것은 당연할 터, UI를 짜증나게 만드는 농담 요소들을 넣는 것을 주체할 수 없었다.

†이스터에그가 하나 있다. 힌트) 野人時代.

- [9] 산업번역혁명, “한영 글자당 효율.” [Online]. Available: <https://rebtion.net/learnfree/?pageid=7&mod=document&uid=508>
- [10] Wall Su, 한국말은 끝까지 들어봐야 하죠. [Online Video]. Available: https://www.youtube.com/watch?v=j7MXqZ5_WG0,