

현대영어교육학회 2020 동계 Workshop

현대영어교육학회(MEESO) 2020 동계WORKSHOP

주제: 코퍼스분석을 위한 Coh-Metrix 프로그램의
실제적 활용방안

일시: 2020년1월10일(금) 13:00 -16:00

장소: 고려대학교문과대학132호

주관: 현대영어교육학회(Modern English Education
Society),
고려대학교영미문화연구소(Institute of British
and American Studies, Korea University)



현대영어교육학회 2020 동계 Workshop

[순서]

12:30-13:00 등록

13:00-13:10 개회사

13:10-13:30 연구윤리특강 (김현진교수)

13:30-14:40 Coh-Metrix 프로그램의 기초 및 관련 연구
(전문기교수)

14:40-14:50 휴식

14:50-16:00 Coh-Metrix 프로그램의 실제적 활용
(전문기교수)

2020 동계 Workshop 준비위원회

김현진(청주교육대학교), 이희경(고려대학교)
이재근(안산대학교), 안수진(인천대학교)
김윤정(경남대학교)

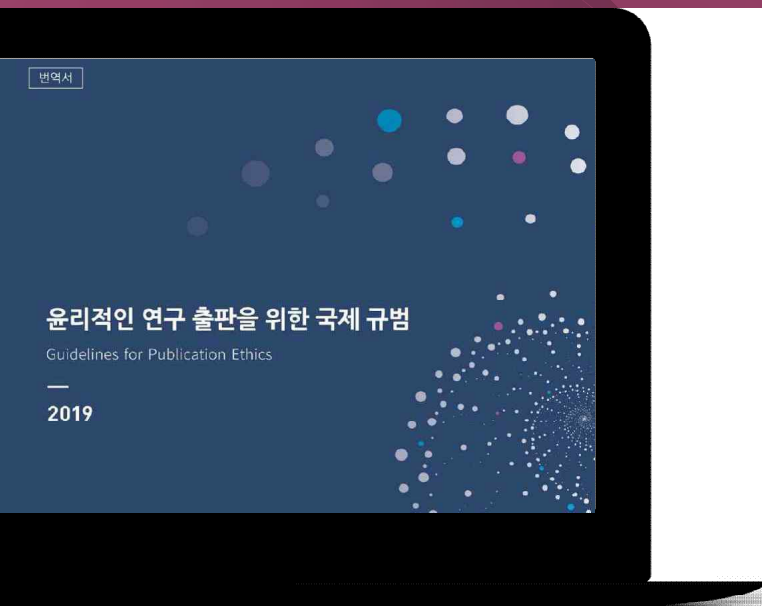
현대영어교육학회 2020 동계 Workshop



2020
현대영어교육학회
동계 WORKSHOP
윤리교육특강

책임 있는 연구 출판

발표자: 김현진
청주교육대학교



번역서

윤리적인 연구 출판을 위한 국제 규범
Guidelines for Publication Ethics
—
2019

발표자: 김현진
청주교육대학교

개발된
0.7.22-24
with Wager

현대영어교육학회 2020 동계 Workshop

개요

발표자료 출처:
책임있는 연구 출판: 저자를 위한 국제 표준

- 윤리적이고 책임 있는 연구 수행
- 명확, 정직하고 조작 없는 결과 제시
- 다른 연구자가 확인 가능하도록 명확한 연구 방법 기술
- 독창적이고 출판된 적이 없는 원고
- 저자는 투고/게재된 작업에 공동책임

개인의 기여를 명확히 표시 연구비 출처와 관련 이해상충을 공개

건전성 및 신뢰성 (Soundness and reliability)	정직성(Honesty)	균형성(Balance)
독창성(Originality)	투명성(Transparency)	적절한 저자됨과 감사문 (Appropriate authorship and acknowledgement)
책무와 책임 (Accountability and responsibility)	전문가 심사 및 출판 규약 준수 (Adherence to peer review and publication conventions)	인간 또는 동물과 관련된 연구의 책임 있는 보고 (Responsible reporting of research involving humans or animals)

1. 건전성 및 신뢰성(Soundness and reliability)

**투고 이후 심사 및 출판 과정
내내 연락 가능해야,
편집위원회의 연락에 바로
반응할 수 있어야**

1.1 보고된 연구는 **윤리적이고 책임 있는 방식으로 수행되어야** 하며 모든 관련 법규를 준수해야 한다. [싱가포르 연구 진실성에 관한 성명서, www.singaporestatement.org]

1.2 보고된 연구는 건전하고 신중하게 수행되어야 한다.

1.3 연구자는 **적절한 데이터 분석 및 표현 방법을 사용해야** 한다(그리고 필요한 경우 이에 대한 전문가의 조언을 구하고 따라야 한다).

1.4 저자는 자신의 연구와 출판물의 내용에 대해 공동 책임을 져야한다. 연구자는 연구의 모든 단계를 주의 깊게 확인하여 **방법과 결과가 정확하게 보고되도록 보장해야** 한다. 저자는 **계산, 데이터 제시, 초고/투고 원고 및 교정본을 주의 깊게 확인해야** 한다.

2. 정직성(Honesty)

2.1 연구자는 **결과를 명확하고 정직하게** 그리고 **위조, 변조 또는 부적절한 데이터 조작없이** 제시해야 한다. 연구 이미지(예 : 현미경 사진, 엑스레이, 전기영동 젤 사진)를 오도하는 방식으로 수정해서는 안 된다.

2.2 연구자는 **연구 방법을 명확하고** 모호하지 않게 기술하여 연구 결과가 **다른 연구자들에 의해 확인될 수 있도록** 노력해야 한다. 연구자는 해당하는 보고 지침을 따라야 한다. 논문에는 충분한 세부 사항을 제공하여 **실험을 다른 연구자가 반복할 수 있도록** 해야한다.

2.3 연구 보고는 완전해야 한다. 부작용, 불일치 또는 설명할 수 없는 소견이나 저자 또는 연구비 지원자의 가설 또는 해석을 뒷받침하지 않는 결과를 생략해서는 안 된다.

현대영어교육학회 2020 동계 Workshop

2.4 연구비 지원자 및 후원자는 자신의 제품이나 입장을 지지하지 않는 조사 결과의 공개를 거부할 수 없어야 한다. 연구자는 연구 후원자가 결과의 공개를 거부하거나 통제할 수 있도록 허용하는 계약을 맺어서는 안 된다(예를 들어 보안상의 이유로 인해 정부가 기밀로 처리한 예외적인 상황을 제외).

2.5 저자가 투고, 승인 또는 출판된 작업에서 오류를 발견한 경우 즉시 편집인에게 이를 알려야 한다. 저자는 편집인과 협력하여 필요한 경우 정정 또는 철회 기사를 게재해야 한다.

2.6 저자는 인용과 출처에서 다른 사람들의 저작물을 정확하게 표시해야 한다.

2.7 저자는 다른 출판물에 인용된 문헌을 읽지 않은 경우 해당 참고문헌을 그대로 복사해서는 안 된다.

3. 균형성(Balance)

철저한 문헌 연구 및 선행 연구의 중요성

3.1 새로운 연구 결과는 이전 연구의 맥락에서 제시되어야 한다. 타인의 작업은 공정하게 표현되어야 한다. 기존 연구의 학술적 검토 및 종합은 완전하고, 균형 있어야 하며, 제안된 가설 또는 해석을 뒷받침하는지 여부와 관계없이 그 결과를 포함해야 한다. 단일한 견해나 논증을 제시하는 논설 혹은 의견은 학술적 종설과 명확하게 구별되어야 한다.

3.2 연구의 한계가 논문에 다루어져야 한다.

서론/결론에 명시

4. 독창성(Originality)

**복수의 학술지에
중복 투고 금지**

**가능한 재인용(as
cited in ...) 하지
않도록**

4.1 저자는 출판 요구 사항을 준수하여 투고된 작업이 독창적이며 다른 언어로 출판되지 않도록 해야 한다. 편집인이 공동 출판에 동의하지 않는 한, 하나의 논문을 하나 이상의 출판사에 동시에 투고해서는 안 된다. 논문이 공동 출판되면 이 사실을 독자에게 분명히 알려야 한다.

4.2 적용 가능한 저작권법 및 규약을 따라야 한다. 저작권 자료(예 : 표, 그림 또는 광범위한 인용문)는 적절한 허가 및 승인을 받은 경우에만 복제해야 한다.

4.3 다른 연구자와 저자 본인의 관련 선행 연구 및 출판물을 적절하게 인정하고 참조해야 한다. 가능하다면 일차 문헌(primary literature)을 인용해야 한다.

**학위 논문을 학술지에 출판하는 경우 투고 시
편집위원회에 알리고, 허가 받으면 이를 논문에
명시해야**

4.4 다른 연구자에 의한 데이터, 텍스트, 그림 또는 개념은 적절하게 인정되어야 하며, 저자 자신의 것처럼 제시되어서는 안 된다. 다른 연구자의 출판물에서 직접 가져온 원문은 적절한 출처와 함께 인용 부호로 표시해야 한다.

4.5 저자는 이전에 연구 결과가 다른 곳에 게재되었거나 혹은 단일 데이터 세트에 대한 여러 보고 또는 분석을 다른 곳에서 발표하기 위해 고려 중인 경우 편집인에게 이를 알려야 한다. 저자는 관련 출판물 또는 다른 학술지에 투고된 원고의 사본을 제공해야 한다.

4.6 단일 연구 사업에서 발생한 여러 출판물은 명확하게 식별되어야 하며 일차 논문을 참조해야 한다. 번역 및 다른 독자층을 위한 각색은 명확하게 식별되어야 하며, 원 출처를 인정해야 하며, 관련 저작권 협약 및 허가 요구 사항을 존중해야 한다. 확신이 서지 않을 경우 저작물을 재출판하기 전에 원본 출판사의 허가를 받아야 한다.

5. 투명성(Transparency)

5.1 직접 및 간접적 재정 지원, 장비 또는 기자재의 공급 및 기타 지원(예 : 전문가 통계 또는 작문 지원)을 포함한 모든 연구비 지원 출처가 공개되어야 한다.

5.2 저자는 연구 설계, 수행, 분석, 해석 및 보고에 대한 연구비 지원 기관(들) 또는 후원자(있는 경우)의 역할을 공개해야 한다.

5.3 저자는 연구 결과의 해석에 영향을 줄 가능성이 있을 것으로 생각되거나, 혹은 편집인, 심사자 또는 독자가 합리적으로 알고 싶어 할 것으로 생각되는 관련 재정적, 비재정적 이해 및 관계를 공개해야 한다. 여기에는 편집인이 자신의 학술지에 자신의 연구를 게시하는 경우와 같이 학술지에 대한 모든 관계도 포함된다. 또한, 저자는 사적인 이익 공개에 대한 학술지와 기관의 요구 사항을 따라야 한다.

6. 적절한 저자됨과 감사문(Appropriate authorship and acknowledgement)

6.1 연구 문헌은 어떠한 사실의 발견을 기록할 뿐만 아니라 발견한 사람이 누구인지도 기록한다. 따라서 연구 출판물의 저자됨은 물론 연구와 보고에 대한 **개인의 기여**를 정확히 반영해야 한다.

- 저자 vs 감사**
- 주저자
 - 제1저자
 - 제2저자
 - 공동저자
 - 교신저자

6.2 주요한 기여자들을 저자로 기재하고, 연구나 출판에 대한 기여도가 낮거나 혹은 순수하게 기술적인 기여만 제공한 이들을 감사문에 기재하는 경우, **저자됨과 감사문의 기준을 연구 시작부터 합의해야** 한다. 이상적으로는 특정 분야의 저자됨 기준에 연구 기관, 전문 학회 및 연구비 지원 기관이 동의하여 게시하고 일관되게 적용해야 한다. 학술지 편집인은 해당 분야에서 받아들여지고 있는 적합한 저자 기준을 게시하고 홍보해야 하지만, 저자됨의 분쟁에서 판결을 제시할 것으로 기대해서는 안 된다. **저자됨의 정확한 귀속에 대한 책임은 소속 기관의 지침 하에 일하는 저자 자신들에게** 있다. 연구 기관은 저자됨과 인정에 대한 공정하고 타당한 표준을 장려하고 지지해야 한다. 필요한 경우 기관은 저자됨의 분쟁을 판결해야 하며 정당한 절차를 따를 것을 보증해야 한다.

**주저자/제1저자/교신 저자가 아닌
저자들도 투고 논문, 최종 출판본을
확인해야**

6.4 모든 저자들은 저자 명단에 오를 것에 동의해야 하며 투고 및 승인된 출판물의 판본을 확인해야 한다. 저자 목록에 대한 변경 사항은 목록에서 삭제된 사람을 포함하여 모든 저자가 승인해야 한다. 교신저자는 편집인과 다른 저자들 간의 연락 지점 역할을 해야 하며, **공동 저자들에게 정보를 제공하고 이들을 출판물에 대한 주요 결정에 참여시켜야** 한다(예 : 심사자 의견에 대한 응답).

6.5 저자는 실제로 연구에 관여하지 않았거나 이를 지지한 적이 없는 개인이 연구에 기여하거나 이를 지지하는 것처럼 보이기 위해 오해의 소지가 있도록 감사문을 사용해서는 안 된다.

7. 책무와 책임(Accountability and responsibility)

**주저자/제1저자/교신 저자가 아닌
저자들도 출판된 논문에 책임이
있음**

7.1 모든 저자는 보고된 작업을 읽고 익숙해져 있어야 하며, 출판물이 이 지침에 명시된 원칙을 준수하는지 확인해야 한다. 대부분의 경우 저자들은 연구의 진실성과 보고에 대해 공동 책임을 질 것으로 기대된다. 그러나 만약 저자들이 연구와 보고의 특정 측면에 대해서만 책임을 진다면, 이는 출판물에 명시되어야 한다.

7.2 저자는 출판 후 오류나 누락을 발견한 경우 즉시 편집인과 이를 정정할 수 있도록 노력해야 한다.

**연구윤리규정에 공동 저자 각각의
역할 명시되어야**

7.3 저자는 재료, 시약, 소프트웨어 또는 데이터 세트를 요청하는 다른 연구자가 이를 이용할 수 있도록 관련 규약, 요구 사항 및 규정을 준수해야 한다. 연구자, 기관 및 연구비 지원 기관은 이러한 요청을 처리하기 위한 명확한 정책을 가지고 있어야 한다. 또한, 저자는 관련 학술지 표준을 따라야 한다. 이를 적절히 인정해 줄 것을 기대할 수 있으나, 연구자는 자료 공유를 조건으로 저자됨을 요구해서는 안 된다.

7.4 저자들은 출판 후 논평과 게재된 독자 통신에 적절하게 응답해야 한다. 저자들은 독자의 질문에 대답하도록 노력해야 하며, 필요한 경우 해명이나 추가 정보를 제공해야 한다.

8. 전문가 심사 및 출판 규약 준수(Adherence to peer review and publication conventions)

복수의 학술지에 중복 투고 금지

신중한 연구 결과 해석 및 결론 도출

8.1 저자들은 출판사의 요구 사항에 따라 하나의 원고를 하나 이상의 출판사에 동시에 제출해서는 안 된다.

8.2 저자들은 조건부 승인을 받은 뒤 원고의 심사를 취소하거나 심사자 의견에 응답하지 않기로 한 경우 이를 편집인에게 알려야 한다.

8.3 저자들은 전문적이며 시의적절하게 심사자의 의견에 응답해야 한다.

8.4 저자들은 출판사의 언론 보도 유예(press embargo) 요구를 존중해야 하며 학술지에 게재가 승인된(하지만 아직 출판되지 않은) 경우 일반적으로 연구 결과가 언론에 보도되는 것을 허용해서는 안 된다. 저자들과 소속 기관은 출판물과 관련된 미디어 활동(보도 자료 및 기자회견 등)을 조정하기 위해 출판사와 연락하고 협조해야 한다. 보도 자료는 연구 결과를 정확하게 반영해야 하며 연구 결과보다 더 나아간 진술을 포함해서는 안 된다.

9. 인간 또는 동물과 관련된 연구의 책임 있는 보고 (Responsible reporting of research involving humans or animals)

9.1 연구가 시작되기 전에 적절한 승인, 허가 또는 등록이 이루어져야 하며, 보고서에 이에 대한 세부 사항이 제공되어야 한다(예 : 기관 심의위원회, 연구 윤리위원회 승인, 동물 사용에 대한 국가 허가 당국).

9.2 편집인의 요청이 있는 경우 저자는 보고된 연구가 적절한 승인을 받았으며 윤리적으로 수행되었음을 입증해야 한다(예 : 승인서, 허가서, 연구 대상자 동의서 사본).

9.3 일반적으로 연구자는 개인(또는 대리인)의 특별한 동의 없이 연구 과정에서 수집된 식별 가능한 개인 데이터를 게시하거나 공유해서는 안 된다. 연구자들은 현재 많은 학술지를 인터넷을 통해 자유롭게 이용할 수 있다는 것을 기억해야 하며 따라서 의도하지 않은 독자에게 위험을 초래하거나 불쾌하게 할 우려가 있음에 주의를 기울여야 한다(예 : 연구대상자 혹은 그 가족들이 사례, 설명, 이미지 또는 가계도에서 자신임을 인식한 경우).

현대영어교육학회 2020 동계 Workshop

9.4 적절한 통계 분석은 연구 시작 시점에 결정되어야 하며 사전 지정된 결과에 대한 데이터 분석 계획을 준비하고 준수해야 한다. 이차 또는 사후 분석은 일차 분석과 데이터 분석 계획에서 설정한 분석과 구별되어야 한다.

9.5 연구자는 이해에 기여할 수 있는 의미 있는 연구 결과를 모두 발표해야 한다. 특히 모든 임상시험 결과를 발표할 윤리적 책임이 있다. 성공하지 못한 연구 또는 가설을 부정한 실험의 발표는 다른 사람들이 유사한 연구에 시간과 자원을 낭비하는 것을 막을 수 있다.

만약 소규모 연구의 결과와 통계적으로 유의미한 결과에 도달하지 못한 연구 결과를 결합하여 보다 유용한 정보를 생성할 수 있다면(예 : 메타 분석을 이용한) 해당 결과를 게재해야 한다.

책임 있는 연구 출판

건전성 및 신뢰성 (Sounness and reliability)	정직성(Honesty)	균형성(Balance)
독창성(Originality)	투명성(Transparency)	적절한 저자됨과 감사문 (Appropriate authorship and acknowledgement)
책무와 책임(Accountability and responsibility)	전문기 심사 및 출판 규약 준수(Achherence to peer review and publication conventions)	인간 또는 동물과 관련된 연 구의 책임 있는 보고 (Responsible report'ng of research invo ving humans or animals)

2020년 현대영어교육학회 동계 Workshop

코퍼스 분석을 위한 Coh-Metrix 프로그램의 실제적 활용 방안

전문기
영어영문학과
건국대학교

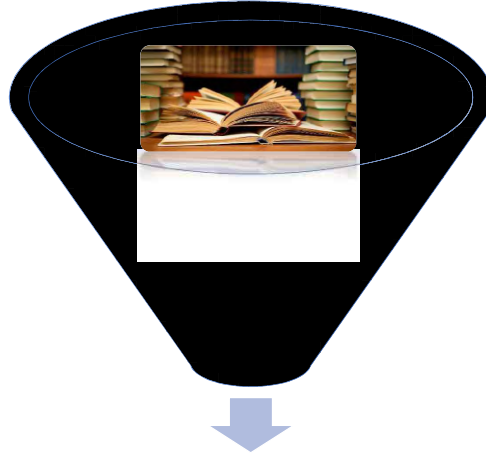


Workshop 개요

Session I

Session II

텍스트의 어떤 측면을 분석할 것인가?



텍스트 정합성 (Text Cohesion)!

텍스트 정합성이란?

- 텍스트 정합성(Text Cohesion)은 텍스트를 구성하고 있는 언어적 요소들에 의해 형성됨
- 구체적으로, 텍스트의 정합성은 텍스트 내에 명확하게 제시된 논항(arguments), 텍스트 지시어(discourse markers), 접속사(connectives) 등의 언어학적 요소들에 의해서 형성됨(Halliday & Hasan, 1976; Graesser, Jeon, Yan, & Cai, 2007; Graesser, Singer, & Trabasso, 1994; Sanders & Matt, 2006).
- 따라서, 정합성은 이러한 언어적 요소들에 기반해 형성된 텍스트에 대한 표상(representation)으로 정의될 수 있음(전문기, 2016)

텍스트 응집성이란?(1/2)

- 텍스트 정합성과 텍스트 응집성(Text Coherence) 여러 학문 분야에서 비슷한 개념으로 종종 혼동되어 사용되고 있지만 본질적으로 개념적 차이가 존재함(전문기, 2016)
- 응집성은 정합성보다 좀 더 확장된 개념으로 정의됨(전문기, 2016)
- 정합성과 달리 응집성은 텍스트 자체에 나타난 언어적 요소 외에 텍스트 이해자의 배경 지식(background knowledge)에 기반해 형성된 텍스트에 대한 표상으로 정의될 수 있음(Graesser, Jeon, Yan, & Cai, 2007)

텍스트 응집성이란?(2/2)

- 다시 말하면, 텍스트의 응집성은 텍스트 기반 정보(text-based information)와 이해자의 배경 지식과의 상호작용 관계에 의해서 형성됨(Graesser, McNamara, Louwerse, & Cai, 2004)
- 부연하면, 텍스트의 정합성은 텍스트 요소들 간의 언어적 연결(linguistic linking)을 의미하고 텍스트의 응집성은 텍스트 내에 존재하는 언어적 요소들과 이해자의 사전 지식(prior knowledge)과의 연결을 지칭함(Taboada, 2004)

왜 정합성이 중요한가?

- 대부분의 언어적 및 심리언어적 측정치들이 텍스트 정합성과 관련이 있음(Graesser, Jeon, Yan, & Cai, 2007; Halliday & Hasan, 1976)
- 텍스트 정합성은 텍스트 이해에 영향을 줌 (Graesser, McNamara, & Louwerse, 2003)
- 텍스트 정합성은 텍스트를 이해하는 능력과 관련 있음(Snow, 2002)
 - i.e., 텍스트 정합성은 교육에 있어서 중요한 관심의 대상임 (McNamara, Louwerse, McCarthy, & Graesser; Jeon, 2008)

McNamara et al. (1996)

- They examined how text cohesion interacts with readers' prior knowledge
- Findings (**reverse cohesion effect**, O'Reilly & McNamara, 2007)
 - High-knowledge readers are more likely to benefit from low cohesion (LC) text
 - Low-knowledge readers are more likely to benefit from high cohesion (HL) text

High-knowledge
readers



Low-knowledge
readers



텍스트의 특성을 어떻게 분석할 것인가?

- 정합성은 텍스트의 이해와 난이도에 영향을 미치는 본질적인 요소이기 때문에 많은 언어심리학자들은 여러 심리학적 방법론과 통제된 실험 등을 통해 텍스트의 정합성에 대해 다양한 측면에서 연구함(Halliday & Hasan, 1976; Kintsch & Van Dijk, 1978; Knott & Dale, 1994; Knott, Sanders, Oberlander, 2001; Long, Wilson, Hurley, & Prat, 2006; Sanders & Noordman, 2000)

정합성의 유형

- Halliday & Hasan (1976)**
 - 참조 정합성(reference cohesion)
 - 대치와 생략 정합성(substitution and ellipsis cohesion)
 - 어휘 정합성(lexical cohesion)
 - 접속 정합성(conjunction cohesion) 정합성
- Graesser, McNamara, & Louwense (2003)**
 - 공통참조(co-reference), 구신정보 단서(given-new cues), 접속사 관계(verb tense), 동사 시제(verb tense), 장면 변화(scene changes), 표제어(headers), 주제문(topic sentences), 구두법(punctuation), 수사구조 표시(signals of rhetorical structure)

정합성의 예

•Halliday & Hasan(1976)

•참조 정합성(reference cohesion)

→ 인칭 참조 정합성

Three blind mice, three blind mice.

See how *they* run! See how *they* run!

정합성의 예

•Halliday & Hasan(1976)

•대치 정합성(substitution cohesion)

My axe is too blunt.

I must get a shaper *one*.

•생략 정합성(ellipsis cohesion)

Joan *brought* some carnations, and Catherine some sweet peas.

정합성의 예

•Halliday & Hasan(1976)

•어휘 정합성(lexical cohesion)

→어휘반복

There was a large *mushroom* growing near her, about the same height as herself; and, when she had looked under it, it occurred to her that she might as well look and see what was on the top of it. She stretched herself up on tiptoe, and peeped over the edge of the *mushroom*,...

→상위 개념어

Henry's bought himself a new Jaguar. He practically lives in the *car*.

정합성의 예

•Halliday & Hasan(1976)

•접속 정합성(conjunction cohesion)

For the whole day he climbed up the steep mountainside, almost without stopping.

And in all this time he met no one. (부가 접속사)

For the whole day he climbed up the steep mountainside, almost without stopping.

Yet he was hardly aware of being tired. (반의 접속사)

For the whole day he climbed up the steep mountainside, almost without stopping.

So by night time the valley was far below him. (인과 접속사)

For the whole day he climbed up the steep mountainside, almost without stopping.

Then, as dusk fell, he sat down to rest. (시간 접속사)

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•공통참조(co-reference)

He was eating an *apple*.
The color of the *apple* was green.

She made a cake for her friends.
It was very tasty.

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•직시(deictic markers)

He visited San Francisco last month.
He had a wonderful time *there*.

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•구신정보(given-new cues)

The jokes Horace tells are awful.

구정보: Horace tells jokes.

신정보: Those jokes are awful.

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•접속사 관계

유형	접속사 예시
부가 접속사	and, also, moreover
시간 접속사	then, when, before, after, during, while
인과 접속사	because, consequently, as a result
의도 접속사	in order to, by means of
반의 접속사	but, although, however
논리 접속사	therefore, so

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•동사시제

- 텍스트 내에 제시된 개개 사건들의 시간적 발생 순서는 텍스트의 정합성 형성에 중요한 영향을 미침
- 동사의 시제 유형(현재, 과거, 미래)과 시간을 나타내는 단서(예를 들면, currently, later, in the future) 등에 의존해서 텍스트에 대한 정합성을 형성

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•장면변화

- 소설과 같은 이야기 텍스트(narrative text)에서 소설 속의 등장 인물들이 활동하는 공간적 장소와 장면의 변화는 텍스트의 정합성 형성에 기여함
- 작가는 특정 공간이나 장면 변화를 지칭하는 다양한 유형의 담화 지시어(예를 들면, back at the ranch)를 사용함으로써 정합적인 이야기를 만들어 냄

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•표제어

- 표제어는 독자들이 텍스트를 읽을 때 그들의 하향식 처리(top-down processing)를 촉진시킬 수 있는 언어적 요소임
- 표제어는 텍스트의 정합성과 응집성 형성에 모두 기여할 수 있음

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•주제문

- 단락의 주요 주제나 생각을 포함하고 있는 주제문은 그 단락의 첫 번째 문장으로 제시되는 경향이 있음
- 텍스트 내에 주제문이 명확하게 제시되면 텍스트의 정합성이 증가하여 결과적으로 그 텍스트에 대한 이해도 증가함

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•구두법

- 텍스트 작성에 적용되는 여러 유형의 구두점(예를 들면, 마침표, 느낌표, 물음표, 쉼표 등)은 텍스트의 정합성 관계를 표현하는데 기여함
- 예를 들면, 대시(dash) 기호(-)는 텍스트 내에 제시된 어떤 개념을 부연 설명할 때 사용될 수 있음
- 이처럼 텍스트의 정합성은 텍스트 내에 사용된 다양한 구두점들에 의해서 형성될 수 있음

정합성의 예

•Graesser, McNamara, & Louwerse(2003)

•수사구조 표시

- 텍스트의 정합성은 텍스트 내에 제시된 다양한 유형의 수사구조들에 의해 형성될 수 있음

유형	수사구조 지시어
비교-대비 구조	whereas, on the one hand, on the other hand
절차 구조	the first step is, the second step is

현대영어교육학회 2020 동계 Workshop

정합성의 유형		
수준	정합성(응집성) 유형(from 전문기, 2016)	
어휘 수준	H&H의 분류	대치와 생략 정합성 어휘 정합성
문장 수준	H&H의 분류	참조 정합성 접속 정합성
	GM&L의 분류	공통참조 구신정보 단서 접속사 관계 동사 시제 구두법
텍스트 수준	GM&L의 분류	장면 변화 표제어 주제문 수사구조 표시

왜 컴퓨터 시스템이 필요한가?

- Need to Analyze **longer texts** such as textbooks, newspaper articles (Van Oostendorp, Otero, & Campanario, 2002), and tutorial dialogs (Jeon, 2008)
 - The manipulation of text variables in the previous studies was possible because sentence pairs, or very small text fragments were used
 - However, when research questions call for longer texts, it becomes more difficult to keep track of the various sources of cohesion and language by hand

왜 컴퓨터 시스템이 필요한가?

•The lack of a **one-stop, computational linguistic department store** (McNamara,

Louwerse, McCarthy, & Graesser, 2010)

- Although there are several computational linguistic databases and techniques available to analyze text, they are distributed across various sources (e.g., **CELEX**, Baayen, Piepenbrock, & van Rijn, 1993; **LSA**, Landauer & Dumais, 1997; **MRC Psycholinguistic database**, Coltheart, 1981; **syntactic parsing**, Charniak, 2000; **WordNet**, Fellbaum, 1998)

대안은?

- 텍스트(빅데이터) 분석 및 평가에 적극적으로 컴퓨터 시스템 활용
- 텍스트 분석 시스템 개발
 - Coh-Metrix(영어 텍스트 정합성 분석)
 - TEES(영어 텍스트 평가 시스템)
 - Auto-Kohesion(한국어 텍스트 정합성 분석)

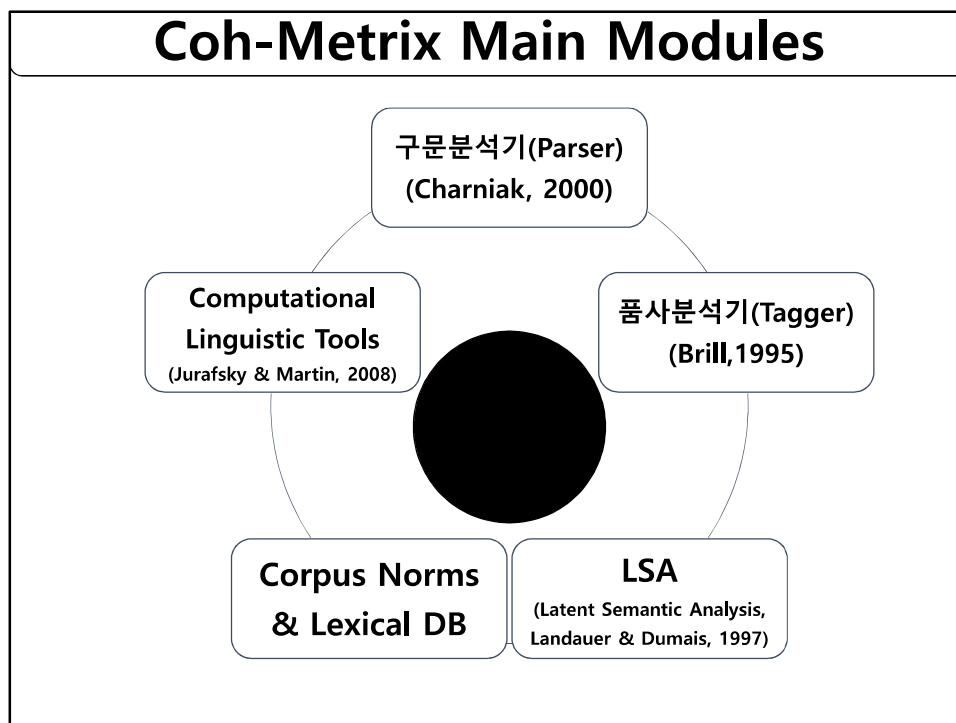
Coh-Metrix

- A **Web-based** automated language analysis tool (Graesser, Jeon, Cai, & Yan, 2007; Jeon, 2008)
- The only computer tool free to the public which was designed to analyze language and textbooks with a rich set of linguistic and psycholinguistic measures

Coh-Metrix

- Analyzes text and discourse on many different dimensions of cohesion, readability, and language
 - Provides 108 measures on public Web site (Coh-Metrix 3.0) (공개용)
 - Provides over 100 measures in PC version (비공개용)
- Put emphasis primarily on deeper levels of meaning and cohesion, unlike standard readability formulas (e.g., Flesch-Kincaid Grade Level, Flesch Reading Ease score)

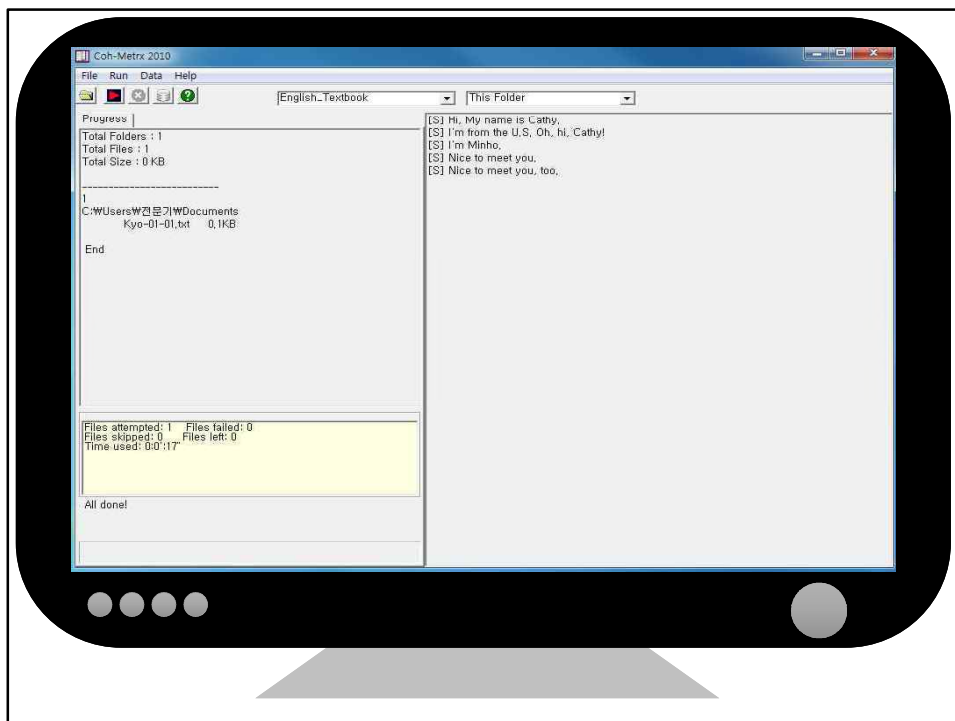
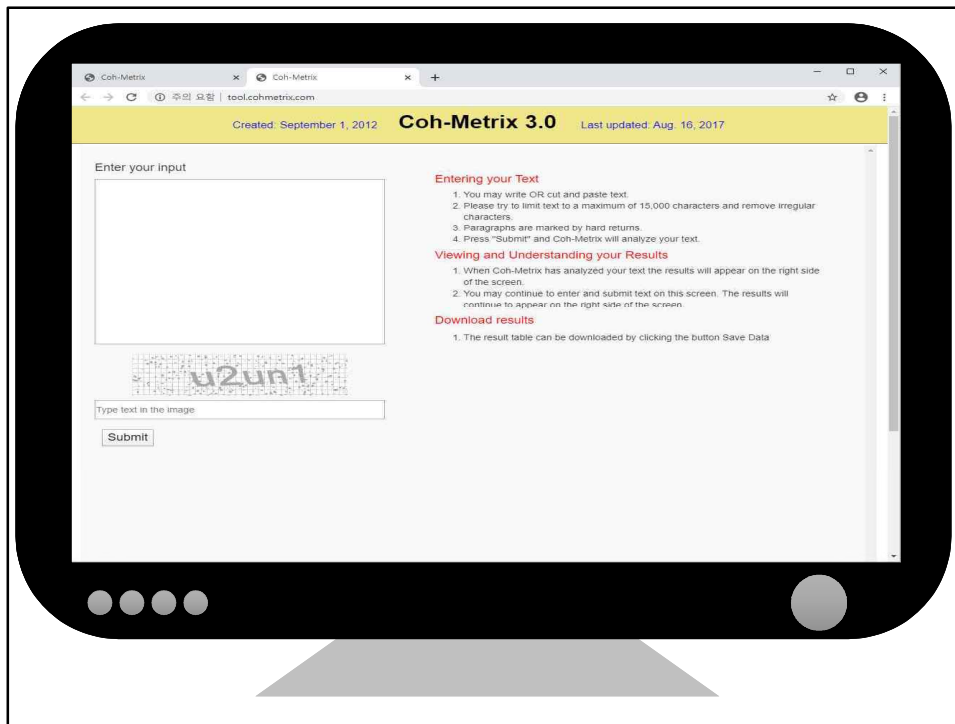
현대영어교육학회 2020 동계 Workshop



Coh-Metrix 시스템의 신뢰도, 타당성 평가

- 전문기 & 최윤희 (2019)
- 이해수 & 전문기 (2019)
- 전문기 (2011, 2014, 2015)
- 김소정 & 전문기 (2016)
- 임인재 & 전문기 (2013)
- 김지은 & 전문기 (2013)
- 전문기 & 임인재 (2009; 2010)
- Jeon (2014)
- So & Jeon (2017)
- Kim & Jeon (2017)
- Kim, Kim, & Jeon (2014)
- Kacewicz, Pennebaker, Davis, Jeon & Graesser (2014)
- Hartley & Jeon (2010)
- Jeon, Graesser, & Cai (2008)
- Jeon & Azevedo (2007; 2008)
- Azevedo & Jeon (2007)
- Jeon, Graesser, McNamara, Louwerse, & Cai (2006)
- Graesser, Jeon, Yang, & Cai (2007)
- Graesser, Jeon, Cai, & McNamara (2008)
- Jeon & Graesser (2006; 2008)
- ...

현대영어교육학회 2020 동계 Workshop



Coh-Metrix 주요 측정치

- 기초산출치(Descriptive Measures)
- 참조적 정합성(Referential Cohesion)
- 의미적 정합성(Semantic Cohesion, LSA)
- 어휘다양성(Lexical Diversity)
- 접속사(Connectives)
- 상황모델(Situation Model)
- 통사적 복잡성(Syntactic Complexity)
- 어휘정보(Word Information)
- 가독성 점수(Readability Scores)

기초산출치(Descriptive Measures)

- 단어 수(The number of words)
 - *DESWC/READNW*
- 문장 수(The number of sentences)
 - *DESSC/READNS*
- 단락 수(The number of paragraphs)
 - *DESPC/READNP*
- 단어길이(Word length)
 - *DESWL_{sy}/READASW*
- 문장길이(Sentence length)
 - *DESSL/READASL*

참조적 정합성(Referential cohesion)

- Co-reference cohesion occurs when a noun, pronoun, or noun phrase refers to another constituent in text or discourse
- Many empirical studies have demonstrated the effect of co-reference cohesion on text and discourse comprehension (Haviland & Clark, 1974; Manelis & Yekovich, 1976; Cirilo, 1981)
- **Higher co-reference cohesion indicates easier understanding of text**

참조적 정합성(Referential cohesion)

- **인접된 문장 쌍에 대한 논항 반복(Argument overlap for adjacent sentences (local cohesion)):** $CRFAO_1/CRFBA_1um$
 - indicates the proportion scores of adjacent sentence pairs in text or discourse which contain common arguments (e.g., nouns, pronouns, noun phrases)
 - Jane has an *apple*.
 - The color of the *apple* is red.
- **전체 문장 쌍에 대한 논항 반복(Argument overlap for all sentences (global cohesion)):** $CRFAO_a/CRFBAaum$
 - indicates the proportion scores for all sentence pairs which contain common nouns, pronouns, or noun phrases

의미적 정합성(Semantic Cohesion, LSA)

• Latent Semantic Analysis (LSA)

- A computational and statistical method that represents the semantic and conceptual meanings of text and discourse based on a large number of corpora (Hu, Cai, Wiemer-Hastings, Graesser, & McNamara, 2007; Jeon, 2008; Landauer, 2007)
- LSA uses singular value decomposition (SVD) to extract approximately 100-500 semantic dimensions from a large word-by-document co-occurrence matrix made from original corpora
- In these semantic dimensions (i.e., a knowledge space) extracted from the word-by-document co-occurrence matrix, a text component (i.e., word, sentence, paragraph, or section) is represented as a vector of values
- The semantic or conceptual similarity between any two text components is computed as a cosine value (functionally varying from -1 to 1) between the two vectors
- **Higher LSA cosines indicate higher semantic and conceptual similarities**

의미적 정합성(Semantic Cohesion)

- **LSA cosine for adjacent sentences (local cohesion): $LSASS1/LSAassa$**
 - This is the mean LSA cosine scores for the adjacent sentences in the text
 - These LSA cosine scores indicate the semantic and conceptual similarities between the adjacent sentences
- **LSA cosine for all sentences**
- **(global cohesion): $LSASSp/LSApssa$**
 - This is the mean LSA cosine scores for all the sentences in the text
 - These LSA cosine scores indicate the semantic and conceptual similarities among all the adjacent sentences

어휘 다양성(Lexical Diversity)

- **Type-token ratio**
 - The word *type* indicates each unique word in a text, whereas the word *token* indicates all instances of each unique word
 - For example, if a word *computer* appears five times in a text, the type of the word *computer* is 1, whereas the token of the *computer* is 5
- **Type-Token Ratio = Number of Type / Number of Token**
- A high type-token ratio of a text indicates that people have much difficulty understanding the text because it contains many unique words to be processed

어휘 다양성(Lexical Diversity)

- **Type-token ratio for all words**
 - $LDTTR_a$
- **Type-token ratio for content words**
 - $LDTTR_c / TYPTOK_c$
- **MTLD (McCarty & Jarvis, 2010):** the mean length of sequential word strings in a text that maintain a given TTR Value
 - $LDMTLD_a / LEXDIVTD$
- **vocd-D (McCarty & Jarvis, 2010):** is calculated through a computational procedure that fits TTR random samples with ideal TTR curves
 - $LDVOCDA / LEXDIVVD$

접속사(Connectives)

- Connectives are important linguistic markers to signal the cohesion and coherence relations (Halliday & Hasan, 1976; Jeon, 2008)
- Coh-Metrix provides several types of connective incidence scores
- *An incidence score indicates the number of occurrences of a particular category per 1,000 words*

접속사(Connectives)

- **Additive connectives** (e.g., also, and, moreover)
 - *CNCAdd/CONADDi*
- **Temporal connectives** (e.g., after, before, when)
 - *CNCTemp/CONTEMPI*
- **Causal connectives** (e.g., because, so, therefore)
 - *CNCCaus/CONCAUSi*
- **Adversative and contrastive connectives** (e.g., although, whereas)
 - *CNCADC/CONADVCONi*
- **Logical connectives** (e.g., and, or)
 - *CNCLogic/CONLOGi*
- **Positive connectives** (e.g., also, moreover)
 - *CNCPos*
- **Negative connectives** (e.g., however, but)
 - *CNCNeg*
- **All connectives**
 - *CNCAII/CONi*

상황모형(Situation Model)

- **Event-Indexing Model** (Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998)
- **Time (Temporal cohesion): deleted**
- **Space (Spatial cohesion): deleted**
- **Protagonist: deleted**
- **Intentionality (Intentional cohesion)**
 - *SMINTE_r/INTEC*
- **Causality (Causal cohesion)**
 - *SMCAUS_r/CAUSC*

통사적 복잡성(Syntactic Complexity)

- Syntactic complexity indicates how difficult it is to analyze the syntactic structures of text or discourse (Graesser et al., 2004)
- Sentences with complex syntactic structures are very difficult to understand because these sentences are structurally dense and have many embedded linguistic components
- Coh-Metrix computes the syntactic complexity based on Charniak Parser (2000)

통사적 복잡성(Syntactic Complexity)

- **A noun phrase (NP) density score: $SYMNP/SYMNP$**
 - The mean number of modifiers per NP
 - The types of these modifiers per NP include adjectives, adverbs, or determiners which qualify head nouns
 - For example, a noun phrase *the very pretty flower* contains three modifiers (the, very, and pretty) and one head noun (flower)
 - Sentences with many modifiers per NP indicate that the sentences are difficult to be understood because of the volume of content in the NP constituents
- **The mean number of words before main verb of main clause in sentences: $SYNLE/SYNLE$**
 - A sentence with many words before the main verb of main clause indicates that people have to hold many words in working memory before they get to the main idea of the sentence that is typically affiliated with the main verb

어휘정보(Word Information)

- The word information in Coh-Metrix includes the **MRC Psycholinguistics Database** (Coltheart, 1981) properties of words, with values ranging from 100 to 700
- MRC contains 150,837 words and provides information about 26 different linguistic properties of these words
- **구체성(Concreteness): $WRDCNCc/WRDCacwm$**
 - How concrete a words is, on the basis of human ratings
- **심상성(Imageability): $WRDIMGc/WRDIacwm$**
 - How easy it is to construct a mental image of the word in one's mind, according to human ratings
- **습득나이(Age of acquisition)**
 - This is the score of the age of acquisition norms (Gilhooly & Logie, 1980) multiplied by 100 to produce a range from 100 to 700
 - Captures the fact that some words appear in children's language earlier than others

현대영어교육학회 2020 동계 Workshop

어휘정보(Word Information)

- CELEX DB에 기반해 어휘빈도 측정치 산출
- 내용어에 대한 어휘빈도(Word frequency for content words): *WRDFRQc/FRCLacwm*
- 모든 어휘들에 대한 어휘빈도(Log frequency for all words):
WRDFRQa/FRCLaewm

어휘정보(Word Information)

- 1인칭 대명사(단수): *WRDPRP1s*
- 1인칭 대명사(복수): *WRDPRP1p*
- 2인칭 대명사: *WRDPRP2/PRO2i*
- 3인칭 대명사(단수): *WRDPRP3s*
- 3인칭 대명사(복수): *WRDPRP3p*
- 대명사 발생점수: *WRDPRO/DENPRPi*
- 명사 발생점수: *WRDNOUN*
- 동사밀도 발생점수: *WRDVERB*
- 형용사밀도 발생점수: *WRDADJ*
- 부사밀도 발생점수: *WRDADV*

현대영어교육학회 2020 동계 Workshop

가독성 점수(Readability Scores)

- **The Flesch Reading Ease score: *RDFRE/READFRE***
 - Ranges from 0 to 100
 - The average of the Flesch Reading Ease score is usually located between 60 and 70
 - A higher Flesch Reading Ease score indicates easier reading
- **Flesch Reading Ease score = $206.835 - (1.015 \times ASL) - (84.6 \times ASW)$**
 - ASL (Average Sentence Length) = The number of words divided by the number of sentences
 - ASW (Average number of Syllables per Word) = The number of syllables divided by the number of words

가독성 점수(Readability Scores)

- **The Flesch-Kincaid Grade level: *RDFKGL/READFKGL***
 - Ranges from 0 to 12
 - Converts the Flesch Reading Ease score into a U.S. grade-school level (i.e., 1 – 12th)
 - A higher Flesch-Kincaid Grade level indicates more difficult reading
- **Flesch-Kincaid Grade level = $(.39 \times ASL) + (11.8 \times ASW) - 15.59$**

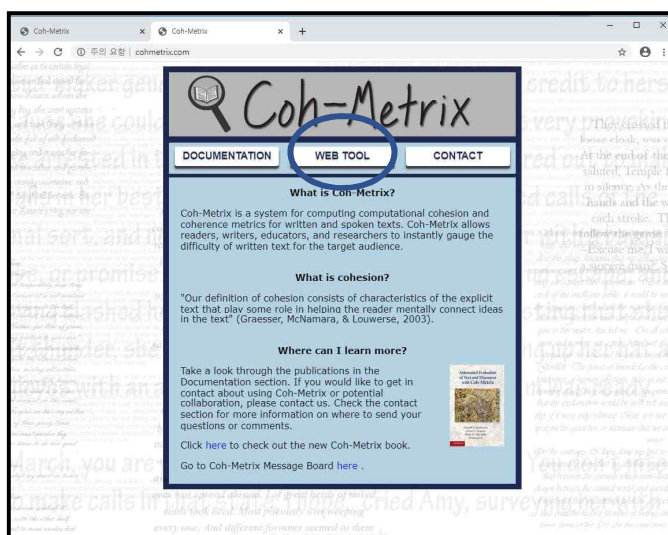
현대영어교육학회 2020 동계 Workshop

가독성 점수(Readability Scores)

- L2 Readability: *RDL2/L2*
 - Based on the subset of Bormuth's (1971) corpus of 32 academic reading texts used by Greenfield (1999) to develop the Miyazaki EFL readability score (Crossley, Allen, & McNamara, 2011)
- $$\text{R2 Readability} = -45.032 + (52.230 * \text{CRFCWO1}) + (61.306 * \text{SYNSTRUT}) + (22.205 * \text{WRDFRQmc})$$
- CRFCWO1: content word overlap
 - SYNSTRUT: sentence syntactic similarity
 - WRDFRQmc: log word frequency for content word

Coh-Matrix 실습

<http://cohmetrix.com/>

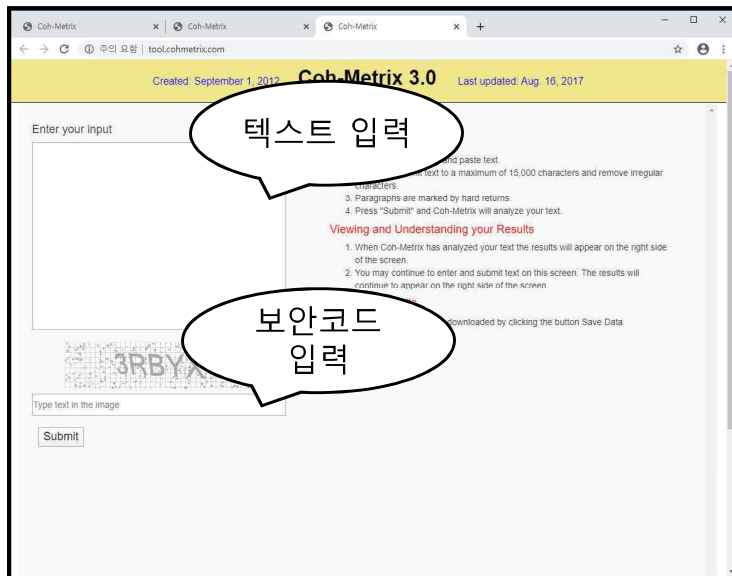


현대영어교육학회 2020 동계 Workshop

Coh-Matrix 실습



Coh-Matrix 실습



현대영어교육학회 2020 동계 Workshop

Which one is more difficult to read(or understand)?

TEXT 1

The mole is a small furry animal that lives underground. It moves around in tunnels that it digs. The mole has changed its life in four ways because it lives underground.

First, the mole has short, solid legs that are good for digging tunnels. This helps the mole dig tunnels so it can hunt for worms to eat. The mole also digs a home tunnel where it sleeps.

Second, good hearing is important so the mole can tell when other animals are in its tunnels. The mole's ears are tiny because larger ears would get in the way when digging.

Third, the mole's eyes are very small and they can "see" only light and dark. This is because moles live underground where it is dark, so it is not important for them to see.

Finally, because it is dark underground, the mole has a special way of figuring out where it is going. When it is going forward, the mole feels its way with its nose and the hairs on its front feet. When the mole is going backward, its tail tells it which way to go.

TEXT 2

Every three to five years along the western coast of South America, the Pacific Ocean's current abruptly changes. This sudden change is called El Nino. The name means "the Child Jesus." It occurs right after Christmas. When El Nino happens, two important climate changes occur. The Pacific coastal waters warm up. In addition, the normal amount of rainfall significantly increases.

Peru suffers the most from the effects of El Nino. The warmer water kills millions of fish. Thousands of birds that feed on the fish also die. And, there are landslides and severe flooding in agricultural regions.

1973 was an El Nino year, and the difficulties Peru faces were especially serious that year. Over-fishing had already reduced the once abundant fish supply. El Nino further reduced to a very small percentage of its former level. Peru's fishing industry has not returned to the supply levels that existed before 1973. Since that time, El Nino continues to severely affect the country. Its fishing and agricultural industries are still in the process of recovery.

2020년 현대영어교육학회 동계 Workshop

코퍼스 분석을 위한 Coh-Metrix 프로그램의 실제적 활용 방안

감사합니다!

전문기
영어영문학과
건국대학교
mjeon1@konkuk.ac.kr

