

언어는 기본문자 위에서 정의한다.

언어를 정의하려면 문자(symbol)들의 집합 기본문자(alphabet, vocabulary; 어휘)를 먼저 정의하여야 한다.

| | |
|---------|--|
| 예) 이진수 | {0, 1} |
| 십진수 | {0, 1, ..., 9} |
| 영어 | {a, b, ..., z} |
| 한글 | {ㄱ, ㄴ, ..., ㅎ, ㅣ, ㅏ, ㅑ, ..., ㅗ} + α^1 |
| Text 파일 | unicode |
| 영어문장 | 영어사전에 나온 단어 |

문자(symbol)들의 집합 기본문자를 Σ^2 라고 하자. 언어(language)는 기본문자 Σ 위에서 (over) 정의(universe of discourse)한다. 문자가 나열되면 문자열(string³)이 된다. 기본문자 {0, 1}에서 정의된 이진수 언어에서 1, 110, 00110등이 문자열의 예이다.

문자열 중에는 특정 언어(language)에 맞는 문자열이 있고 그렇지 않은 문자열이 있다. 그래서 언어(language)를 문자열의 **집합**으로 정의한다. 예를 들어 ㅎㅏㄱㄱㅓ 문자열은 한글 어법에 맞는 한글 문자열이지만, 스ㅋㅌ르 문자열은 현대한글 어법에 **맞지 않으므로** 한글 문자열이 아니다. 이 경우 한글 **어법에 맞는** 한국어 **문자열**만을 모은 **집합**을 한글(언어; language)로 정의할 수 있다.

문자열의 **길이**는 그 문자열에 들어 있는 문자의 개수이다. 예를 들어 |1| = 1이고 |110| = 3, |00110| = 5이다. 문자열의 **연결**(concatenation, \cdot)을 정의할 수 있다. 예를 들어 문자열 school과 boy를 **연결**하면 school \cdot boy =⁴ schoolboy이다. 길이 6개짜리 문자열 school과 길이 3개짜리 문자열 boy를 **연결**해서 길이 9개짜리 새로운 문자열 schoolboy를 만든다.

길이가 0인 **빈 문자열**(empty string)을 생각하고, ϵ^5 이 **빈 문자열**을 나타낸다고 하자. **빈 문자열**은 어떤 문자열과 **연결**하여도 그 문자열을 바꾸지 않는다. 즉 ϵ 과 school을 연결하거나 school과 ϵ 을 연결하여도, 결과는 그냥 school일 뿐이다⁶). **빈 문자열** ϵ 을 **연결 연산**에서 **항등원**(identity element)⁷이라고 부르고 $|\epsilon| = 0$ 이고 기본문자 Σ 에 관계없이 항상 ϵ 으로 표시한다.

문자열의 전체집합(universe)은 무엇일까? 기본문자 $\Sigma = \{0, 1\}$ 인 이진수 언어를 생각하자. 이진수 문자열은 길이가 1인 0, 1이 있고⁸ 길이 2인 00, 01, 10, 11과 길이 3인 000, 001,

1) α 에 관하여는 2장 한글 모아쓰기 기계 참조

2) 그리스어 대문자 Sigma이다. 소문자는 σ 이다.

3) 문자열 = a sequence of symbols. 고등학교 때 배운 수열은 숫자들의 나열임을 기억하자.

4) 이렇게 쓰는 것을 **그대로쓰기**(juxtaposed)라고 한다.

5) ϵ 은 그리스 문자 epsilon이다. 어떤 교과서에서는 그리스 문자 Lambda인 Λ 나 λ 를 쓰기도 한다.

6) $\epsilon \cdot \text{school} = \text{school} \cdot \epsilon = \text{school}$.

7) $+$ 연산에서 항등원은 0이고, \times 연산에서 항등원은 1이다.

8) 문자는 길이가 1인 문자열의 일종이다.

..., 111등이 있다. 이것을 각각 $\Sigma^1, \Sigma^2, \Sigma^3$ 으로 정의한다면 이진수의 전 집합은 $\Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \dots$ 으로 표시할 수 있을 것이다. 여기에 길이가 0인 빈 문자열 ϵ 도 포함하여, $\{\epsilon\}$ 을 Σ^0 으로 정의하면, 이진수 전체집합은 빈 문자열을 포함하여 $\Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \dots$ 으로 정의할 수 있다.

(정의) 기본문자를 Σ 라고 하자. 기본문자에 n 번 반복 $\Sigma^n (n \geq 0)$ 을 아래로 정의한다.

Basis $\Sigma^0 \stackrel{B}{=} \{\epsilon\}.$

Recursion $\Sigma^n \stackrel{R}{=} \Sigma \cdot \Sigma^{n-1}, \text{ 단 } n \geq 1.$

(예) $\{0, 1\}^2 \stackrel{R}{=} \{0, 1\} \cdot \{0, 1\} \stackrel{R}{=} \{0, 1\} \cdot \{0, 1\} \cdot \{0, 1\}^0 \stackrel{B}{=} \{0, 1\} \cdot \{0, 1\} \cdot \{\epsilon\}$
 $= \{0, 1\} \cdot \{0, 1\} = \{0 \cdot 0, 0 \cdot 1, 1 \cdot 0, 1 \cdot 1\} = \{00, 01, 10, 11\}.$

(정의) 기본문자 Σ 의 반복합 Σ^\dagger 과 Σ^* 을 아래로 정의한다.

$$\Sigma^\dagger \stackrel{B}{=} \bigcup_{i \in N_1} \Sigma^i = \Sigma^1 \cup \Sigma^2 \cup \Sigma^3 \cup \dots \quad \text{단 } N_1 \stackrel{B}{=} \{1, 2, 3, \dots\}.$$

$$\Sigma^* \stackrel{B}{=} \bigcup_{i \in N_0} \Sigma^i = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots \quad \text{단 } N_0 \stackrel{B}{=} \{0, 1, 2, \dots\}.$$

Σ^\dagger 는 길이가 1이상인 모든 문자열에 집합, Σ^* 은 빈 문자열 ϵ 도 포함하여 길이가 0이상인 모든 문자열의 집합을 나타내며, Σ^* 를 문자열의 전체집합(universe)으로 본다.

네 가지 기본용어(terminologies)

언어는 기본문자 Σ 를 우선 정의하여야 한다. 이 때 언어이론의 네 가지 기본용어인 (1) 문자는 $a \in \Sigma$ 로 (2) 기본문자는 Σ 로, (3) 문자열은 $x \in \Sigma^*$ 로 (4) 언어는 $L \subseteq \Sigma^*$ 로 쓴다.

| | | |
|-----------------------|--|--|
| | 원소 | 집합 |
| 길이 1 | ⁽¹⁾ 문자(symbol), $a \in \Sigma$ | ⁽²⁾ 기본문자(vocabulary), Σ |
| 길이 0이상 ¹⁰⁾ | ⁽³⁾ 문자열(string), $x \in \Sigma^*$ | ⁽⁴⁾ 언어(language), $L \subseteq \Sigma^* (L \subseteq 2^{\Sigma^*})$ |

기본문자 Σ 에서 정의한 언어 $L \subseteq \Sigma^*$ 에 대하여, 문자열 $x \in \Sigma^*$ 가 언어 L 에 속하면($x \in L$) 문장(sentence), 그렇지 않으면 ($x \notin L$) 문장이 아니(non-sentence)라고 부른다.

연결에 (\cdot) 정의역과 치역을 언어 $L \subseteq \Sigma^*$ 로 일반화 한다.

(정의) $\cdot : 2^{\Sigma^*} \times 2^{\Sigma^*} \rightarrow 2^{\Sigma^*}.$

$L, S \in 2^{\Sigma^*}$ 라 하자.

$$L \cdot S = \{x \cdot y \in \Sigma^* | x \in L, y \in S\}^{11}.$$

연결기호 \cdot 는 생략하기도 한다(juxtaposed).

원소문제(membership problem)

기본문자 Σ 에서 정의한 언어 $L \subseteq \Sigma^*$ 에 대하여, 문자열 $x \in \Sigma^*$ 가 문장인가 아닌가를 정하는 문제를 언어 L 의 원소문제라고 부른다.

9) Σ^\dagger 는 Σ dagger라고 읽고, Σ^* 는 Σ star라고 읽는다.

10) 문자는 길이가 1인 문자열이다.

11) $L^* = L^\dagger \cup \{\epsilon\}$. $L^\dagger = L^* - \{\epsilon\}$, if $\epsilon \notin L$.

$$L: \Sigma^* \rightarrow \{\mathbf{true}, \mathbf{false}\}$$

$$L(x) = \mathbf{if } x \in L \rightarrow \mathbf{true} \mid x \notin L \rightarrow \mathbf{false} \mathbf{fi.}$$

결정문제(Decision problem, 문제)

집합 D 의 원소 $d \in D$ 에 대하여 $P(d)$ 의 결과가 **true** 또는 **false**인 함수 P 를 문제라 한다.

$$P: D \rightarrow \{\mathbf{true}, \mathbf{false}\}$$

언어의 원소문제는 정의역(domain) D 가 Σ^* 인 결정문제의 일종이다¹²⁾.

12) 강의의 후반부에 가면, 언어와 문제를 구분하지 않을 것이다.