<u>סיכום חומר ההרצאות באוטומטים ושפות פורמליות</u>

- מילה היא סדרה **<u>סופית</u>** של אותיות. אין חסם סופי על אורך המילים ב Σ^*, אבל כל מילה היא באורך סופי. ε היא המילה הריקה מעל כל א"ב.
- שרשור של שפה עם השפה הריקה ϕ, זה ϕ. שרשור שפה עם ε זה השפה המקורית.
- החזקה ה-0 של שפה היא ε.
- בגרף של אוטומט סופי דטרמיניסטי מכל מצב יוצאת **<u>בדיוק</u>** קשת אחת מסומנת σ **<u>לכל</u>** $\sigma\in\Sigma$.
- **<u>הגדרה פורמלית של אוטומט A</u>**: $A=(\Sigma,Q,q_0,F,\delta)$ משמאל לימין: א"ב, קבוצת מצבים, מצב התחלתי יחיד השייך ל-Q, קבוצת מצבים מקבלים המוכלת ב-Q, ופונקצית מעברים.
- $\hat{\delta}:Q\times\Sigma^*\to Q$ הרחבת פונקצית המעברים δ. כלומר תיאור המעבר ממצב כלשהו עם קלט של מילה למצב אחר.
- תכונה חשובה של חישוב באוטומט: $\delta(q,xy)=\delta(\delta(q,x),y)$
- **<u>השפה של אוטומט</u>** – הגדרה פורמלית: $L(A)=\{x\in\Sigma^*\mid\hat{\delta}(q_0,x)\in F\}$ אוסף כל המילים אשר מעבירות את האוטומט מהמצב ההתחלתי למצב מקבל כלשהו.
- השפה של מצב q: $Lq=\{w\mid\delta(q_0,w)=q\}$ אוסף כל המילים המעבירות את האוטומט מהמצב ההתחלתי למצב q.
- שפה $L\subseteq\Sigma^*$ היא **<u>שפה רגולרית</u>** אמ"מ קיים אוטומט סופי דטרמיניסטי A שמקבל את השפה L. יתכן שקיים יותר מאוטומט אחד כזה.
- <u>שפות רגולריות טריוויאליות</u>: ϕ, ε, Σ^*, כל שפה סופית.
- **<u>לא</u>** כל תת קבוצה של שפה רגולרית היא שפה רגולרית.
- <u>תכונות סגור של שפות רגולריות</u>: שפות רגולריות סגורות תחת: איחוד סופי, חיתוך סופי, משלים, חיסור, שרשור, איטרציה, הומומורפיזם, הומומורפיזם הפוך וחלוקה (ראה בהמשך).
- שפות רגולריות <u>אינן</u> סגורות תחת איחוד וחיתוך <u>אינסופיים</u>.
- הגדרת אוטומט סופי אי-דטרמיניסטי זהה להגדרת אוטומט דטרמיניסטי (החמישייה) מלבד הגדרת δ. באוטומט סופי א"ד δ מוגדרת כך: $\delta:Q\times\Sigma\to 2^Q$ כלומר מכל מצב ניתן ללכת ל-0 או יותר מצבים עם אותו קלט. ההרחבה למילים: $\hat{\delta}:Q\times\Sigma^*\to 2^Q$.
- <u>הגדרת שפת אוטומט סופי א"ד</u>: $L(A)=\{x\in\Sigma^*\mid\hat{\delta}(q_0,x)\cap F\neq\varnothing\}$ אוסף כל המילים כך שקיים עבורן חישוב המגיע למצב מקבל.
- לכל אוטומט סופי אי דטרמיניסטי עם מסעי ε קיים אוטומט סופי אי דטרמיניסטי ללא מסעי ε ואוטומט סופי דטרמיניסטי שקול.
- <u>שפת הביטויים הרגולריים</u> – הגדרה אינדוקטיבית: בסיס: $\phi\in R,\ \varepsilon\in R,\ \Sigma\subseteq R$. צעד: אם $r1,r2\in R$ אז: $(r1+r2)\in R$, $(r1r2)\in R$, ו- $(r1)^*\in R$.
- כל ביטוי רגולרי $r\in R$ מגדיר שפה רגולרית המסומנת $L[|r|]$. כל שפה רגולרית ניתן לבטא בעזרת ביטוי רגולרי.
- משפט Kleene: משפחת השפות הרגולריות היא משפחת השפות הקטנה ביותר המכילה את כל השפות הסופיות וסגורה לפעולות רגולריות (איחוד, שרשור ואיטרציה).
- **<u>למת הניפוח לשפות רגולריות</u>**: **<u>לכל</u>** שפה רגולרית L, **<u>קיים</u>** מספר טבעי n, כך ש**<u>לכל</u>** מילה $z\in L$ כך ש $|z|\geq n$ **<u>קיים</u>** פירוק $z=uvw$ המקיים:
  1. $|uv|\leq n$
  2. $|v|\geq 1$
  3. $uv^iw\in L$ לכל $i\geq 0$. *הערה:* יתכן $u=\varepsilon$ או $w=\varepsilon$.

- קיום למת הניפוח היא תנאי <u>הכרחי</u> לרגולריות, אבל לא תנאי <u>מספיק</u>.
- הגדרת הומומורפיזם: יהיו Σ, Γ שני אלפבתים. הומומורפיזם הוא העתקה $h : \Sigma \to \Gamma^*$. כל הומומורפיזם ניתן להרחבה באופן יחיד עבור מילים: $h : \Sigma^* \to \Gamma^*$ כך ש $h(xy)=h(x)h(y)$. *הערה:* $h(\varepsilon) = \varepsilon$.
- הגדרת הומומורפיזם הפוך: $h^{-1} : 2^{\Gamma^*} \to 2^{\Sigma^*}$ מוגדר ע"י $h^{-1}(L) = \{x \in \Sigma^* \mid h(x) \in L\} \subseteq \Sigma^*$ לכל $L \subseteq \Gamma^*$. כלומר עבור מילה מסוימת ב Γ^* ההומומורפיזם ההפוך שלה הוא כל המילים ב Σ^* שעוברות למילה הזאת בהומומורפיזם הרגיל.
- תכונות הומומורפיזם: 1. $h(L1L2)=h(L1)h(L2)$. 2. $h(\cup L_i) = \cup(h(L_i))$.
- הערה: הכיוון השני לסגירות שפות רגולריות להומומורפיזם <u>לא מתקיים</u>. כלומר $h(L)$ רגולרית לא אומר בהכרח ש- L רגולרית.
- הגדרת פעולות $L1, L2 \subseteq \Sigma^*$:

חלוקה מימין - $L_1 / L_2 = \{x \in \Sigma^* \mid \exists y \in L_2, xy \in L_1\}$ כלומר כל הרישות של מילים ב-L1 כך שהמחרוזת המשלימה להם נמצאת ב L2.

חלוקה משמאל - $L_2 \setminus L_1 = \{x \in \Sigma^* \mid \exists y \in L_2, yx \in L_1\}$ כלומר כל הסיפות של מילים ב-L1 כך שהמחרוזת המשלימה להן נמצאת ב L2.

- אם L1 רגולרית, ו-L2 שפה כלשהי, אזי L1/L2 ו L2\L1 רגולריות גם כן.
- אוטומט סופי דטרמיניסטי מקבל מילה כלשהי אמ"מ הוא מקבל מילה $w \in \Sigma^*$ כך ש $|w|<|Q|$.
- לאוטומט סופי דטרמיניסטי A, $L(A)$ אינסופית אמ"מ קיימת מילה $w \in L(A)$ כך ש $|Q| \le |w| < 2|Q|$.
- <u>**הגדרה פורמלית של דקדוק G**</u>: $G=(V,T,P,S)$ משמאל לימין – קבוצה סופית לא ריקה של משתנים, קבוצה סופית לא ריקה של טרמינלים, משתנה תחילי מתוך V, קבוצה סופית של כללי שכתוב.
- הגדרת <u>שפה הנוצרת ע"י דקדוק G</u>: $L(G)=\{x \in T^* \mid S \Rightarrow^* x\}$ כלומר אוסף כל המילים הטרמינליות הנגזרות מ-S.
- הגדרת <u>**דקדוק חסר הקשר**</u>: דקדוק בו כל כלל הוא מהצורה: $A \to \alpha$ כאשר $\alpha \in (V \cup T)^*$.
- הגדרת <u>**דקדוק רגולרי**</u>: דקדוק בו כל כלל הוא מהצורה: $S \to \varepsilon$, $A \to a$, $A \to aB$, $A \to Ba$ כאשר $A,B \in V$ ו- $a \in T$. *הערה:* דקדוק רגולרי הוא בפרט חסר הקשר.
- <u>שקילות דקדוקים לינאריים ושפות רגולריות</u>: לכל שפה רגולרית L קיים דקדוק לינארי ימני G כך ש $L=L(G)$, ולכל דקדוק לינארי ימני G קיים אוטומט סופי M כך ש $L(G)=L(M)$.
- שקילות דקדוקים לינאריים: לכל דקדוק לינארי ימני קיים דקדוק לינארי שמאלי השקול לו ולהפך.
- הגדרה: $\alpha \in (V \cup T)^*$ היא <u>**מילת חזית**</u> של עץ גזירה בדקדוק חסר הקשר G אמ"מ $S \Rightarrow^* \alpha$.
- אם T1 הוא תת עץ גזירה של T2 ו $\alpha 1$ היא מילת חזית של T1, ו- $\alpha 2$ היא מילת חזית של T2 אזי $\alpha 1$ היא תת מילה של $\alpha 2$.
- צעד גזירה יקרא <u>**שמאלי ביותר**</u> (<u>**ימני ביותר**</u>) אם משוכתב בו המשתנה השמאלי ביותר (ימני ביותר) בתבנית. גזירה תקרא שמאלית ביותר (ימנית ביותר) אמ"מ כל צעדי הגזירה בה הם שמאליים ביותר (ימניים ביותר).
- דקדוק חסר הקשר G הוא <u>**רב משמעי**</u> אמ"מ קיימת מילה $w \in L(G)$ אשר לה לפחות 2 עצי גזירה שונים. אחרת G <u>**חד משמעי**</u>.
- רב משמעיות היא תכונה של דקדוקים לא של שפות. לפעמים (אך לא תמיד) קיים דקדוק חד משמעי שקול.
- דקדוק יקרא <u>**פשוט**</u> אמ"מ אין בו משתנים מיותרים (משתנים שאינם משתתפים בגזירה טרמינלית), כללי ε, וכללי יחידה.
- דקדוק הוא <u>**מהצורה הנורמלית של חומסקי**</u> אמ"מ כל הכללים בו מהצורה: $A,B,C \in V,\ a \in T,\ A \to BC \mid a$
- דקדוק הוא <u>**מהצורה הנורמלית של גרייבך**</u> אמ"מ כל הכללים בו מהצורה: $A \in V,\ \alpha \in V^*,\ a \in T,\ A \to a\alpha$.
- לכל דקדוק ח"ה G, אם $\varepsilon \notin L(G)$, אזי קיים דקדוק שקול בצורה נורמלית חומסקי, וקיים דקדוק שקול בצורה נורמלית גרייבך.
- לכל מילה $w \in L(G)$ כאשר G דקדוק ח"ה בצנ"ח, אם גובה עץ הגזירה של w לא גדול מ-i אזי $|w| \le 2^{i-1}$, ואם הגובה הוא בדיוק n אזי $n=2|w|-1$.

- **<u>למת הניפוח עבור שפות חסרות הקשר</u>**: **לכל** שפה חסרת הקשר L, **קיים** מספר טבעי n, כך **שלכל** $z\in L$ כך ש $|z|\geq n$, **קיים** פירוק $z=uvwxy$ כך ש:
  1. $|vwx|\leq n$
  2. $|vx|\geq 1$
  3. $uv^iwx^iy\in L$ לכל $i\geq 0$

  כמו כן, $n\leq 2^{|v|}$ כאשר $|v|$ מספר המשתנים בדקדוק צנ"ח ל $L-\{\varepsilon\}$.
- הגדרת **<u>אוטומט מחסנית</u>**: $M=(Q,\Sigma,\Gamma,\delta,q_0,\text{-|},F)$. משמאל לימין: קבוצת מצבים סופית ולא ריקה, א"ב הקלט, א"ב המחסנית, פונקציית מעברים, מצב תחילי שייך ל-Q, אות מיוחדת מתוך Γ המייצגת את תחילת המחסנית, קבוצת מצבים מקבלים מוכל ב-Q.

  פונקציית המעברים מוגדרת: $\delta : Q\times(\Sigma\cup\{\varepsilon\})\times\Gamma\rightarrow 2^{Q\times\Gamma^*}$
- אוטומט מחסנית מקבל מילה אמ"מ הוא במצב מקבל או המחסנית ריקה. שני אופני הקבלה שקולים.
- תיאור רגעי של אוטומט מחסנית הוא $id=(q,w,\alpha)$ כאשר q הוא המצב הנוכחי, $w\in\Sigma^*$ יתרת הקלט, ו- $\alpha\in\Gamma^*$ תוכן המחסנית.
- ההגדרה הבסיסית של אוטומט מחסנית היא של אוטומט אי-דטרמיניסטי. בכדי שאוטומט מחסנית יהיה דטרמיניסטי שני התנאים הבאים חייבים להתקיים:
  1. לכל $q\in Q$, $a\in\Sigma\cup\{\varepsilon\}$, $Z\in\Gamma$: $|\delta(q,a,Z)|\leq 1$ כלומר יש לכל היותר אפשרות אחת לעבור ממצב כלשהו.
  2. לכל $q\in Q$, $Z\in\Gamma$: אם $\delta(q,\varepsilon,Z)\neq\varnothing$ אז לכל $\sigma\in\Sigma$: $\delta(q,\sigma,Z)=\varnothing$. לא יתכן גם מעבר ε וגם מעבר רגיל ממצב כלשהו.
- לכל שפה חסרת הקשר L קיים אוטומט מחסנית M כך ש $L=L(M)$ כאשר אופן הקבלה הוא מחסנית ריקה.
- מחלקת השפות חסרות ההקשר סגורה תחת: איחוד, שרשור, איטרציה, הצבה ח"ה (ובפרט הומומורפיזם), וסגורה לחיתוך רק עם שפה רגולרית.
- מחלקת השפות חסרות ההקשר אינן סגורות תחת חיתוך עם שפה לא רגולרית, ומשלים.