הרצאה 9

1 דקדוק חסר הקשר – Γ_2

דקדוק $G=(T,V,P,S)$ ייקרא חסר הקשר (ח"ה) אם כל כלל גזירה ב-P הוא מהצורה $A\to\alpha$ כאשר $A\in V$, $\alpha\in(V\cup T)^*$.

בנוסף, הכלל היחיד שמקצר הוא $S\to\epsilon$. (משתנים אחרים לא יכולים).

1.1 שפה חסרת הקשר

נאמר ששפה $L\subseteq\Sigma^*$ היא ח"ה אם קיים דקדוק ח"ה G כך ש $L=L(G)$.
(נובע מכאן שגם שפות רגולריות הן ח"ה, אך לא להיפך).

דוגמה לבניית דקדוק לשפה:

תהי $L=(a^lb^mc^md^n : 0\le l\le n,\ 1\le m)$. נבנה לשפה דקדוק ח"ה.

קל לראות ש L אינה רגולרית (בעזרת למת הניפוח לשפות רגולריות. ואינטואיטיבית, כי צריך לספור). נראה ש L ח"ה ע"י בניית דקדוק מתאים. בכללי הגזירה נלך מהחוץ פנימה.

קל לראות ש $T=\{a,b,c,d\}$.

נרצה משתנה שבהינתן a ייצר גם d (כדי לקיים את התנאי $l\le n$). $R\to aRd\mid ad$.

אותו משתנה גם יכול להוסיף עוד d כרצוננו, מבלי תלות באותיות אחרות. $R\to Rd\mid d$.

נרצה משתנה שייצר את b^mc^m של האמצע (ראינו כבר בנייה כזו). $M\to bMc\mid bc$.

וגם, צריך להבטיח שנשתמש פעם אחת לפחות ב-M. כדי שיתקיים $1\le m$. $R\to M$.

הבעיה היא שזה לא **מכריח** אותנו להגיע ל-M. אז נשנה את הכללים שכבר כתבנו – לא נאפשר ל-R לעצור בלי להגיע ל-M. כלומר נוריד את האפשרות לשכתב רק טרמינלים. בסה"כ, נקבל את הכללים:

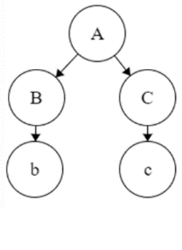
$$P=\{R\to aRd\mid Rd\mid M,\qquad M\to bMc\mid bc\}$$

והדקדוק: $G=(\{a,b,c,d\},\{R,M\},P,R)$.

1.2 עצי גזירה

לסדרת גזירה של w כלשהו בדקדוק נתון G, בד"כ קיימות סדרות גזירה נוספות השונות ממנה רק בסדר של "ניתוח" המשתנים. בד"כ נרצה לחשוב על סדרות כאלה בתור שקולות. למשל, לקומפיילר אין חשיבות לסדר "פיתוח" המשתנים.

דוגמה: $G=(\{A,B,C\},\{b,c\},A,P)$ שבו $P=\{A\to BC, B\to b, C\to c\}$, למילה $w=bc$ יש שתי סדרות גזירה שונות עד כדי סדר הניתוח: $A\to BC\to bC\to bc,\quad A\to BC\to Bc\to bc$. נאמר כי שתי גזירות אלה שקולות.



כדי לתפוס את כל הסדרות השקולות לסדרה נתונה, במובן זה, ניתן לבטא את תהליך הגזירה כ"עץ גזירה". המילה המתקבלת היא רצף העלים משמאל לימין. כאמור, כל סדרת גזירה מגדירה עץ גזירה יחיד, אבל עץ גזירה יכול להגדיר מספר סדרות גזירה שונות בסדר הפיתוח.

הגדרה פורמלית: עץ גזירה עבור דקדוק ח"ה $G=(T,V,P,S)$ הוא עץ סופי, <u>סדור</u> (סדר הבנים חשוב), המקיים:

1. כל צומת מסומנת בסימן מתוך $V\cup T\cup\{\epsilon\}$.
2. השורש יסומן בד"כ S (לפעמים נדבר על תתי עצים, ושם זה לא בהכרח מתקיים).
3. הסימון של כל צומת פנימי שאינו עלה, הוא משתנה מ-V.
4. צומת המסומן ב-ϵ הוא בן יחיד.
5. אם צומת פנימי מסומן ב-A, ולבניו יש סימונים $X_1,X_2,\dots X_t$ (בסדר הזה!), אזי $(A\to X_1X_2\cdots X_t)\in P$. (כלומר A משכתב את $X_1X_2\cdots X_t$).

1.3 חזית של עץ גזירה

הגדרה: המחרוזת המתקבלת ע"י שרשור של כל העלים משמאל לימין הינה **חזית** של עץ הגזירה. אם החזית של עץ הגזירה מורכבת ממילה טרמינלית, נאמר שהעץ הזה הוא **עץ גזירה מלא**.

מה הקשר בין עץ גזירה לבין תבנית פסוקית של דקדוק?

טענה: α היא מילת חזית עץ הגזירה עם שורש A (בדקדוק נתון G) אמ"מ $A \Rightarrow_G^* \alpha$.

ולכן עצי גזירה שימושיים להבנת המילים המתקבלות.

הוכחת הטענה באינדוקציה: כיוון אחד. נראה שאם α הינה חזית של עץ גזירה חוקי ששורשו A, אזי $A \Rightarrow_G^* \alpha$.
באינדוקציה חזקה על מספר הצמתים הפנימיים בעץ:

בסיס: עבור $n = 0$ צמתים פנימיים, כל העץ הוא A. ולכן חזית העץ היא $A = \alpha$ ואכן $A \Rightarrow_G^* \alpha$ כי $A \Rightarrow^0 A$.
(במקרה זה, העץ אינו עץ גזירה מלא, שכן חזית העץ אינה מילה טרמינלית).

יהי $i \in \mathbb{N}^+$. נניח שהטענה נכונה לכל $j < i$ צמתים פנימיים, ונראה עבור i צמתים.

היות ו- $i > 0$, יש ל-A בנים כלשהם. נסמנם $X_1, X_2, \dots X_t$ (בסדר הזה).

מהגדרת עץ הגזירה, בדקדוק G קיים הכלל $A \to X_1 X_2 \cdots X_t$.

נתבונן בתתי העצים של $X_1, X_2, \dots X_t$: ממבנה העץ כולו מתקיים ש $\alpha = \alpha_1 \alpha_2 \cdots \alpha_t$.
(כאשר לכל i, הביטוי α_i הינו החזית של תת העץ הנובע מ X_i).

היות וקיימים הצמתים $X_1, X_2, \dots X_t$ (לפחות אחד), לכל אחד מתתי העצים הללו יש לכל היותר $i - 1$ צמתים פנימיים, ולכן הנ"א חלה עליהם. לכן, בדקדוק G, לכל i קיימות סדרות הגזירה $X_i \Rightarrow^* \alpha_i$. ולכן קיימת הסדרה:

$$A \to X_1 X_2 \cdots X_t \Rightarrow^* \alpha_1 \alpha_2 \cdots \alpha_t = \alpha$$

מסקנה: קבוצת מילות החזית של כל עצי הגזירה המלאים בדקדוק G היא $L(G)$. כי ראינו שכל מילת חזית היא מילה בשפה. וגם שכל מילה בשפה, היא מילת חזית.

1.4 תת עץ גזירה

תת עץ גזירה הוא חלק מהעץ המכיל צומת כלשהו ואת כל צאצאיו.

טענה (לא נוכיח): יהי T_1 תת עץ גזירה של T_2, ותהיינה α_1, α_2 מילות החזית של T_1, T_2 בהתאמה. אזי, α_1 תת מילה של α_2.

2 צורות נורמליות

2.1 פישוט דקדוק

בהינתן דח"ה, לפעמים הוא ייראה מסובך מדי, ונרצה לפשט אותו כדי להוכיח עליו טענה מסויימת. לדוגמה, בשביל למת הניפוח.
הרעיון הכללי של הפישוט יקביל, במקצת, לרעיון הצמצום של אס"ד.

ארבעה אלגוריתמים הנוגעים לפישוט דקדוקים:

1. סילוק משתנים שאינם טרמינליים,
2. סילוק משתנים לא ישיגים,
3. סילוק כללי אפסילון,
4. סילוק כללי יחידה.

(לא נראה אותם בקורס הזה).

שתי צורות נורמליות של דקדוקים:

הצורה הנורמלית של חומסקי: כל כללי הגזירה הם מהצורה $A \to BC$ או $A \to \sigma$.
(באופן טבעי, אין דקדוק רגולרי בצורה הנורמלית של חומסקי).

הצורה הנורמלית של גרייבך: כל כללי הגזירה הם מהצורה $A \to a\beta$. (a היא אות טרמינלית).
כלומר, בכל צעד גוזרים לפחות אות אחת, ואז עוד מחרוזת מעל $V \cup T$.

נתייחס לשפות שלא מכילות ϵ (אחרת מסתבכים עם דקדוקים שבהם $S \to AS$ וכדומה).

משפט: כל שפה חסרת הקשר (ללא אפסילון) ניתנת לייצור ע"י דקדוק בצורה נורמלית. נוכיח לגבי הצורה הנורמלית של חומסקי.
(בצורה של גרייבך לא נשתמש בקורס).

2.1 מעבר לצורה הנורמלית של חומסקי

נתרגם את G המייצר את L לצורה הנורמלית של חומסקי: (לא למדנו את שלבים 1,2).

1. סילוק כללי אפסילון (כלומר, כללים מהצורה $A \to \epsilon$) וסילוק משתנים שאינם ניתנים להשגה.
2. סילוק כללי יחידה ($A \to B$). (זה לא מייצר כללי אפסילון – היזכרו בחלק השלישי של כללי הגזירה מהאלגוריתם ההוא).
3. תרגום כללים מהצורה $A \to X_1 X_2 \cdots X_k$ (כאשר $k > 2$) לצורה חוקית:
   a. אם $X_i = \sigma$, אזי נוסיף משתנה חדש C_σ ואת הכלל $C_\sigma \to \sigma$, ונחליף את X_i ב C_σ. נבצע את זה לכל כלל גזירה ב-P ולכל טרמינל בו.
   b. לכל $A \to B_1 \cdots B_k$ עבור $k \geq 3$: נוסיף משתנים $D_1 \ldots D_{k-2}$ ונחליף את הגזירה בקבוצת הגזירות
$$A \to B_1 D_1, \qquad D_1 \to B_2 D_2$$
$$D_{k-3} \to B_{k-2} D_{k-2}, \qquad D_{k-2} \to B_{k-1} B_k$$

3 למת הניפוח לשפות חסרות הקשר

אם L שפה חסרת הקשר, אזי קיים n כך שלכל מילה $z \in L$ המקיימת $|z| \geq n$ קיים פירוק מהצורה $z = uvwxy$ כאשר:

1. $|vwx| \leq n$
2. $1 \leq |v| + |x|$
3. $\forall i \geq 0 : uv^i wx^i y \in L$

בנוסף, אם G דקדוק ח"ה מהצורה הנורמלית של חומסקי יוצר את $L \setminus \{\epsilon\}$ וגם $|V(G)| = k$,
(כלומר אם יצרנו דקדוק ח"ה לשפה לפי הצורה הנורמלית של חומסקי ויש לנו k משתנים), אזי $n \leq 2^k$.

3.1 הוכחת הלמה – מבט על

נראה שבהינתן מילה z מספיק ארוכה השייכת לשפה, נוכל לפרק אותה לחמישה חלקים $z = uvwxy$, המקיימים (עבור משתנים A, S מהדקדוק):

1. $S \to^* uAy$
2. $A \to^* vAx$
3. $A \to^* w$

למה? כי אם ניתן להגיע למשתנה A, והוא גוזר את עצמו עם תוספת, קיבלנו ניפוח (ופינצ'ור) בדומה למעגל שראינו בלמת הניפוח לשפות רגולריות.

3.2 מסלול בעץ

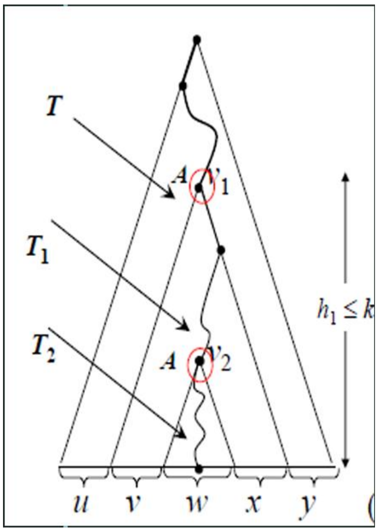
יהי G דקדוק בצורה הנורמלית של חומסקי היוצר את $L \setminus \{\epsilon\}$, ויהי $|V(G)| = k$. תהי $z \in L$.
(אגב, מה עושים עם אפסילון? טכנית, מחקנו מילה מהשפה. זה בסדר כי היא לא ארוכה יותר מ-n).

גובה עץ הגזירה של z הינו לפחות $\log_2(|z|) + 1$. (למה? כי אם נתעלם רגע מהשלב האחרון של הטרמינלים, העץ העמוק ביותר שיכול להתקבל הוא עץ בינארי מלא בן $|z|$ עלים. ולכן גובה העץ הוא לפחות $\log_2(|z|)$. נחזיר את הטרמינלים, ולכן הוספנו עוד 1 לגובה העץ).

יהי $n = 2^k$, ותהי $z \in L$ כך ש $n \leq |z|$. אזי נקבל שגובה עץ הגזירה של z מקיים:

$$k + 1 = \log_2(2^k) + 1 \leq \log_2(|z|) + 1 \leq h$$

כלומר, מסלול ארוך ביותר בעץ מכיל לפחות $k + 2$ צמתים, שהם לפחות $k + 1$ צמתים פנימיים שמסמנים משתנים. (ספירת העומר, שם שם). היות ויש רק k משתנים, לפי עקרון שובך היונים קיים במסלול משתנה המופיע פעמיים, כלומר שני צמתים המכילים את אותו המשתנה. זה ה"מעגל" – אם היינו בשלב מסויים במשתנה כלשהו, עשינו צעדים והגענו שוב לאותו משתנה. וכמובן נוכל לעשות את אותם הצעדים שוב.



במסלול הארוך ביותר הזה, נסתכל מהעלה כלפי מעלה, עד שנגיע למשתנה שמופיע פעמיים – בפעם העליונה, זה יקרה עד הגובה $h_1 \leq k$. (למה? בהמשך).

נסמן את המשתנה הזה ב-A, ואת שני הצמתים עם A נסמן בתור v_1 (הקרוב לשורש), ו-v_2.
נסמן T_1 את תת-העץ המושרש ב-v_1, ו-T_2 את תת-העץ המושרש ב-v_2.

בגלל שאנחנו בצורה הנורמלית של חומסקי, ה-A הראשון משכתב שני משתנים. אם אחד מהם (בה"כ הימני) הוא ה-A השני, אז במקרה הזה x יהיה ריק. אבל אז v יהיה בוודאות לא ריק, אז אנחנו בסדר.

נסמן את מילת החזית של T_2 ב-w. לפי הטענה מתחילת ההרצאה: $A \Rightarrow^* w$.

T_2 הוא תת עץ של T_1 ולכן מילת החזית של T_1 היא vwx, ושוב: $A \Rightarrow^* w$.
ויותר מכך, נוכל לרשום $A \Rightarrow^* vAx$. (נסמן ב **).

לסיום – מילת החזית של כל העץ היא $z = uvxyz$ ולכן $S \Rightarrow^* uAy$. (נסמן *).

במסלול הארוך ביותר הזה, נסתכל מהעלה כלפי מעלה, עד שנגיע למשתנה שמופיע פעמיים – בפעם העליונה. זה יקרה עד הגובה $h_1 \leq k$. למה?

אם המשתנה העליון גבוה יותר, אזי קיים מישהו מתחתיו שגם מופיע פעמיים, וניקח אותו במקום. בנוסף, היות ולקחנו את המסלול הארוך ביותר, אין מסלול המוביל מ-A לעלים שאורכו גבוה יותר. חסם הגובה הה מאד יעיל, בזכות חומסקי קיבלנו שמילת החזית של הA הראשון חסומה באורך 2^k.

3.3 הפירוק

כעת נראה כי הפירוק $z = uvwxy$ מקיים את תנאי הלמה:

1. כיוון שגובה T_1 מקיים $h_1 \leq k$, נקבל כי חזית העץ vwx מקיימת $|vwx| \leq 2^k = n$. פה השתמשנו בכך שבצורה הנורמלית של חומסקי, כל הכללים מהצורה $A \rightarrow BC$.
2. $1 \leq |v| + |x|$. נראה כי לא ייתכן ששניהם ϵ.
בצעד הראשון מ-v_1 לכיוון v_2 הפעלנו כלל מהצורה $A \rightarrow BC$.
אם T_2 נגזר מ-B, אזי בתת העץ C יש לפחות אות אחת (ואז $1 \leq |x|$).
אם T_2 נגזר מ-C, אזי בתת העץ B יש לפחות אות אחת (ואז $1 \leq |v|$).

התנאי השלישי: נוכיח באינדוקציה על i כי $S \Rightarrow^* uv^iAx^iy$.

בסיס: עבור $i = 0$ נקבל $S \Rightarrow^* uAy$. ואכן $S \Rightarrow^* uAy$, כפי שהראנו בבנייה.

נניח כי הטענה מתקיימת עבור $i - 1$, כלומר $S \Rightarrow^* uv^{i-1}Ax^{i-1}y$. נוכיח עבור i:

$$S \Rightarrow^* uv^{i-1}Ax^{i-1}y \Rightarrow^* uv^{i-1}vAxx^{i-1} = uv^iAx^iy$$

כאשר החץ הראשון נובע מהנ"א, והשני מ **.

כדי לסיים את הוכחת הלמה נותר רק להשתמש בעובדה ש $A \Rightarrow^* w$, כפי שהראנו בבנייה (חזית העץ T_2). ונקבל uv^iwx^iy.

4 דוגמאות

4.1 דוגמה 1

תהי $L = \{a^jb^jc^j : j \geq 1\}$ מעל $\Sigma = \{a,b,c\}$. הוכיחו כי L אינה ח"ה.

נב"ש שהיא כן ח"ה, ויהי n הקבוע המובטח בלמה. נבחר: $z = a^nb^nc^n \in L$. מתקיים $|z| = 3n \geq n$. נתבונן בפירוקים המקיימים את תנאי הלמה:

$$a \ldots \ {}_1 \ldots a \ {}_4 \ b \ldots \ {}_2 \ldots b \ {}_5 \ c \ldots \ {}_3 \ldots c$$

המספרים 1-5 מסמנים מיקום אפשרי של vwx.

1. $vwx = a^{|vwx|}$
2. $vwx = b^{|vwx|}$
3. $vwx = c^{|vwx|}$
4. $vwx = a \ldots b$
5. $vwx = b \ldots c$

עבור פירוק 1, נבחר $i = 2$ ונקבל $a^{n+|vx|}b^nc^n \notin L$, בסתירה ללמה. מקרים 2,3 דומים.

עבור פירוק 4 נבחר $i = 0$ ונקבל $a^lb^kv^n \notin L$, שכן $l < n$ או $k < n$. מקרה 5 דומה.

בכל הפירוקים מצאנו i שעבורו המילה המנופחת לא בשפה, ולכן L אינה ח"ה.

4.2 דוגמה 2

נוכיח ש $L = \left\{a^{j^2} : j \geq 1\right\}$ אינה ח"ה:

נב"ש שהיא כן, ויהי n הקבוע המובטח בלמה. נבחר $z = a^{n^2}$. מתקיים $|z| = n^2 \geq n$ ולכן ע"פ הלמה קיים פירוק $z = uvwxy$ המקיים: $|x| + |v| \geq 1$, $|nwx| \geq n$. בכל פירוק: $vwx = a^{|vwx|}$. נסמן: $v = a^t, x = a^k$ (כאשר $t \geq 1 \vee k \geq 1$). עבור $i = 2$ נקבל: $uv^2wx^2y = a^{n^2+t+k}$. מתקיים: $n^2 < n^2 + t + k \leq n^2 + n < (n+1)^2$ (כי $|vwx| \leq n$). כלומר המילה המנופחת לא בשפה, בסתירה ללמה. אז L אינה ח"ה.

אם השפה הייתה: $L = \left\{a^{n^2}b^k : n \geq 1, k \geq 0\right\}$, לא היינו מצליחים. כי אז "האויב" היה בוחר פירוק שבו $vwx \in b^*$, ואז כל ניפוח מנפח רק את ה-b, והמילה הייתה נשארת בשפה.

4.3 דוגמה 3

תהי $L = \{a^nb^n : 1 \leq n\}$, נראה שהיא מקיימת את למת הניפוח לשפות חסרות הקשר:

נבחר n (נגיד בהמשך מה הוא), לכל $z \in L$ כך ש $|z| \geq n$ נגדיר פירוק $z = uvwxy$:

אנחנו צריכים פירוק שעונה על תנאי הלמה, כך שכל ניפוח של v, x ישאיר את המילה בשפה. נבחר: $v = a, x = b, w = \epsilon$. האותיות u, y יהיו הקצוות. נגדיר פורמלית: לשם נוחות, נגדיר $t = |z|/2$. ואז $u = a^{t-1}, y = b^{t-1}$. ואז כל ניפוח מוסיף את אותו מספר ש a, b, והסדר שלהם נשמר (כל ה-a ואז כל ה-b).

צריך שזה יתקיים גם עבור $i = 0$. במקרה הזה, ה-x, v נעלמים ונישאר רק עם u, y. נדרוש שהם יהיו ארוכים מספיק: $1 \leq t - 1$, כלומר $2 \leq t$, כלומר $4 \leq |z|$. וזה מגדיר את ה-n שלנו.

4.4 דוגמה 4

תהי $L = b^*c^*d^* \cup a^* \cdot \{b^jc^jd^j : j \geq 1\}$. נראה שהיא מקיימת את למת הניפוח לשפות חסרות הקשר:

כל מילה מהצורה $b^*c^*d^*$, אפשר לנפח בלי בעיה.

כל מילה מהצורה $a^* \cdot \{b^jc^jd^j : j \geq 1\}$, נבחר $v = a, x = \epsilon$ וננפח. כדי שנוכל לבחור $v = a$, נדרוש $n = 1$.

אם הבחירה של a^* היא ϵ, המילה היא בעצם מהצורה $b^*c^*d^*$.

כלומר, לא הצלחנו להראות שהיא לא ח"ה. בשביל זה נצטרך תכונות סגור.

5 תכונות סגור

כמו שיש תכונות סגור לשפות רגולריות, יש תכונות סגור לשפות ח"ה.

בהינתן L, L_1, L_2 שפות ח"ה, ו-R שפה רגולרית, גם השפות הבאות ח"ה:

1. איחוד: $L_1 \cup L_2$.
2. שרשור: $L_1 \cdot L_2$.
3. איטרציה: L^*.
4. היפוך: L^R.
5. חיתוך עם שפה רגולרית: $L \cap R$.
6. הומומורפיזם: $h(L)$.
7. הומומורפיזם הפוך: $h^{-1}(L)$.
8. הצבה חסרת הקשר.

את תכונות 6-8 לא נסביר ולא נשתמש בהן.

הערה חשובה: יכול להיות מצב שהפעולה הפכה שפה ח"ה לשפה רגולרית. זה תקין, כי כל שפה רגולרית היא גם ח"ה. אפשר לעלות בדרגות ההיררכיה של חומסקי, לא לרדת.

5.1 איחוד

בהינתן L_1, L_2 שפות ח"ה, קיימים להן דקדוקים ח"ה $G_1 = (V_1, T_1, S_1, P_1),\ G_2 = (V_2, T_2, S_2, P_2)$. אזי הדקדוק הבא מתאים לשפה $L_1 \cup L_2$:

$$G_{1\cup 2} = (V_1 \cup V_2 \cup \{S\}, \qquad T_1 \cup T_2, \qquad S, \qquad P_1 \cup P_2 \cup \{S \to S_1 | S_2\})$$

זה בהנחה כי $V_1 \cap V_2 = \emptyset$, וגם $S \notin V_1 \cup V_2$. אחרת צריך להחליף שמות, אבל הרעיון זהה.

הוכחה על רגל אחת: בצעד הגזירה הראשון נבחר האם אנחנו רוצים מילה מ-L_1 או מ-L_2.

הוכחה על שתי רגליים:

כיוון ראשון: תהי $w \in L_1 \cup L_2$, יש להראות כי $w \in L(G_{1\cup 2})$. מתקיים:

$w \in L_1$ ואז $S_1 \Rightarrow^* w$, או $w \in L_2$ ואז $S_2 \Rightarrow^* w$.

כיוון שני: תהי $w \in L(G_{1\cup 2})$. היות וכל המילים הנגזרות ב-$G_{1\cup 2}$ הן מהצורה $S \to S_1 \Rightarrow^* w$ או $S \to S_2 \Rightarrow^* w$, אזי $w \in L_1$ או $w \in L_2$.

5.2 שרשור

בהינתן L_1, L_2 שפות ח"ה, קיימים להן דקדוקים ח"ה $G_1 = (V_1, T_1, S_1, P_1),\ G_2 = (V_2, T_2, S_2, P_2)$. אזי הדקדוק הבא מתאים לשפה $L_1 \cdot L_2$:

$$G_{1\cdot 2} = (V_1 \cup V_2 \cup \{S\}, \qquad T_1 \cup T_2, \qquad S, \qquad P_1 \cup P_2 \cup \{S \to S_1 S_2\})$$

הוכחה על רגל אחת: בצעד הגזירה הראשון נגזור למשתנה ההתחלתי של L_1 וגם למשתנה ההתחלתי של L_2, וכך נקבל מילה מ-L_1 משורשרת עם מילה מ-L_2.

5.3 איטרציה

בהינתן L שפה ח"ה, קיים לה דקדוק ח"ה $G = (V_1, T_1, S_1, P)$. ניצור דקדוק עבור L^*:

$$G_* = (V_1 \cup \{S\}, \qquad T_1, \qquad S, \qquad P \cup \{S \to S_1 S \mid \epsilon\})$$

הוכחה על רגל אחת: בצעד הגזירה הראשון נאפשר לקבל חזרה את S כדי לגזור מילה נוספת מ-L_1.ושיימאס לנו נבחר ב-ϵ ונעצור. וכמובן, אפשר לבחור בו מלכתחילה, שכן איטרציה מכילה את אפסילון.

5.4 אי סגירות לחיתוך

בהינתן L_1, L_2 שפות ח"ה, השפה $L_1 \cap L_2$ לא בהכרח חסרת הקשר.

הוכחה: תהי $L_1 = \{a^n b^n c^m : n, m \geq 1\} \cap \{a^n b^m c^m : n, m \geq 1\}$. כל תת שפה הינה חסרת הקשר (ניתן לבנות דח"ה או אוטומט מחסנית מתאים) אך החיתוך ביניהן הוא $L_\cap = \{a^n b^n c^n : n \geq 1\}$ שזו שפה תלוית הקשר (נכשלת בלמת הניפוח).

מסקנה מזה – אי סגירות למשלים (כי אחרת, הייתה סגירות לחיתוך, לפי כללי דה מורגן).

5.5 היפוך

ניעזר בצורה הנורמלית של חומסקי.

תהי L ח"ה ויהי $G = (V, T, P, S)$ כך ש $L = L(G)$. אזי הדקדוק הבא יוצר את L^R:

$$G^R = (V, T, P', S)$$

$$P' \coloneqq \{A \to CB : A \to BC \in P\} \cup \{A \to \sigma | A \to \sigma \in P\}$$

הוכחה על רגל אחת: הפכנו את סדר הגזירה של כל המשתנים, ולכן כל המילים יוצאות הפוכות.

5.6 חיתוך עם שפה רגולרית

ההוכחה מסתמכת על אוטומט מחסנית, ומשתמשת באוטומט מכפלה.

יהי M_L א"מ עבור L, ויהי A_L אס"ד עבור R. נבנה אוטומט מכפלה ביניהם. נתקדם במקביל בשני האוטומטים. זה טיפה יותר מסובך כי יש מסעי אפסילון (שבהם לא מתקדמים ב-A_L), ויש אי-דטרמיניזם. לכן, נתקדם בא"מ, ושם נגדיר את מצב הבקרה, מ-A_L.

6 תרגילים

הוכיחו/הפריכו בעזרת תכונות סגור ולמת הניפוח: בהינתן $L_1 = \{a^n b^n : n \geq 1\}, L_2 = \{a^{2n} b^{3n} : n \geq 2\}$ שפות ח"ה, גם השפות הבאות ח"ה:

ראשית, נראה שגם L_2 ח"ה: ניצור דקדוק ח"ה עבורה: $S \to aaSbbb \mid \epsilon$.

$L_3 = \{a^n b^n a^{2n} b^{3n} : n \geq 1\}$. לכאורה נראה שרשור של $L_1 L_2$, אבל נשים לב שהדרישה פה קשה יותר: כי זה אותו n. זאת תהיה הפרכה.

$L_4 = \{a^n b^n a^n b^n : n \geq 1\}$. גם פה לכאורה שרשור $L_1 L_1$, אבל זה אותו n. גם הפרכה.

$L_5 = \{a^n b^n a^{2k} b^{3k} : n, k \geq 1\}$. זה אכן שרשור.

$L_6 = \{a^n b^n \cup a^{2n} b^{3n} : n \geq 1\}$. למרות שזה אותו n, זה איחוד. כי כל פעם בוחרים מילה מכאן <u>או</u> מכאן.

$L_7 = \{d^*\} \cup \{a^n b^n c^n : n \geq 1\}$. נב"ש שהיא ח"ה, כלומר $L_7 \cap \{a^* b^* c^*\}$ גם ח"ה (חיתוך עם שפה רגולרית). אבל החיתוך הוא $a^n b^n c^n$, שהיא לא ח"ה (ראינו מקודם).

הפרכה עבור L_3: נראה באמצעות למת הניפוח שהיא לא ח"ה:

נב"ש שהיא ל"ה ויהי n הקבוע המובטח בלמה. נבחר $z = a^n b^n a^{2n} b^{3n} \in L_3$, המילה ארוכה מספיק ולכן קיים פירוק לפי הלמה. אפשרויות לפירוק: $a \ldots {}_1\ a\ {}^5\ b \ldots {}_2\ b\ {}^6\ a \ldots {}_3\ a\ {}^7\ b \ldots {}_4\ b$.

עבור מקרה 1: נבחר $i = 2$ ונקבל: $a^{n+|vx|} b^n a^{2n} b^{3n} \notin L_3$. מקרים 2,3,4 דומים.

עבור מקרה 5 נבחר $i = 0$ ונקבל: $a^{n-r} b^{n-k} a^{2n} b^{3n} \notin L_3$ (כי $1 \leq r$ או $1 \leq k$).

לכל מקרה מצאנו i שעבורו המילה המנופחת לא בשפה, בסתירה ללמת הניפוח, כלומר השפה אינה ח"ה.

הפרכה עבור L_4 עובדת בצורה דומה. השוויון של ה-n מחייב זיכרון שאפילו אוטומט מחסנית לא מספק.