10

נושא 2 - דקדוקי שכתוב

א) הקדמה

עד כה למדנו 3 דרכים להגדרת שפה $L\subseteq\Sigma^*$.

(1) באמצעות הגדרות המוכרות מתורת הקבוצות. לדוגמה: $L=\{a^p : p \text{ ראשוני}\}$. זוהי הגדרה הנוחה ביותר אך אינה עוזרת לחשב האם מילה בשפה או לייצר מילה מהשפה.

(2) באמצעות אוטומט או מחשב* המקבלות מילה ומחזירות כן/לא אם מילה בשפה. *מודל

(3) על ידי תהליך רקורסיבי לייצור מילים מהשפה. דוגמה לכך היא ביטוי רגולרי שניתן לייצר איתם בעזרת אופרטורים ייצור מילים מהשפה שהביטוי הרגולרי מייצג. בחינה שימושית לחישוב האם מילה בשפה.

דקדוקי שכתוב הם מודל הדומה לביטויי רגולרי, באופן זה שניתן לייצר מהם מילים מהשפה שהדקדוק מייצג. אולם להבדיל מביטויים רגולריים, דקדוקים ניתן לכתוב גם עבור שפות שאינן רגולריות.

ב) הגדרה

דקדוק שכתוב G מוגדר באמצעות 4 פרמטרים $G=(V,T,S,P)$:

V - קבוצה סופית של משתנים. נסמן בד"כ באמצעות אותיות גדולות. בהמשך נפרט כיצד להשתמש בהם.

T - קבוצה סופית של טרמינלים. מייצגים את האותיות בשפה שהדקדוק מייצג. ניתן לומר $T=\Sigma$. T ו-V הם קבוצות זרות לחלוטין $V\cap T=\emptyset$. נסמן טרמינלים ע"י בד"כ באמצעות אותיות קטנות.

S - משתנה התחלתי $S\in V$. בכל פעם שיוצרים מילה מתחילים ממשתנה זה בלבד.

P - קבוצה סופית של כללי גזירה, מהצורה $\alpha\to\beta$. אלו כללים שבאמצעותם אפשר להמיר קטעים ממחרוזת נוכחית α אל מחרוזות אחרות β, ובכך לייצר מילים חדשות. α ו-β מוגדרים באופן הבא: $\alpha\in(V\cup T)^*V(V\cup T)^*$, $\beta\in(V\cup T)^*$. כלומר, שניהם יכולים להכיל משתנים וטרמינלים, אך α חייב להכיל לפחות משתנה אחד. מכך ש-β יכול להכיל משתנים נוכל להפעיל כללי גזירה באופן רקורסיבי.

במקרה שיש כמה אפשרויות להמיר את α, נהוג להשתמש בסימון המקוצר $\alpha\to\beta_1|\beta_2|\beta_3|\ldots$.

<u>שפת הדקדוק:</u> נאמר כי מילה $w\in\Sigma^*$ שייכת לשפה של הדקדוק G, המסומנת $L(G)$, אם ישנה איזושהי סדרת גזירה מ-S כך שנקבל את w. סדרה של פעולות גזירה חייבת להתחיל מ-S ולהסתיים ב-T^* ללא משתנים מ-V. נסמן כלל גזירה באמצעות "$\to$" ופעולת גזירה "$\Rightarrow$". שפת הדקדוק G מוגדרת להיות $L(G)=\{w\in T^*: S\Rightarrow^* w\}$.

<u>הערות:</u> (1) מקובל לעתים להציג דקדוק באמצעות קבוצה P בלבד, שהרי מכילה את כל שאר הפרמטרים.

(2) אם ישנו דקדוק G שבו לא נוכל להגיע למילה טרמינלית באמצעות כללי הגזירה, נאמר כי שפת הדקדוק $L(G)=\emptyset$.

דוגמה: שפת הדקדוק $G=(V=\{S\}, T=\{a,b\}, S, P=\{S\to aSb \mid ab\})$ היא השפה $L(G)=\{a^nb^n : n\ge 1\}$.

דוגמה לסדרת גזירה: $S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow a^4b^4 \in L(G)$

ג) היחס $\Rightarrow_G$

נוכל להתייחס לפעולת הגזירה $\Rightarrow$ כאל יחס בינארי מעל $(V\cup T)^*$.

- יהי G דקדוק שכתוב ויהיו $\alpha,\beta\in(V\cup T)^*$, נאמר ש-$(\alpha,\beta)\in\Rightarrow_G$, ובקיצור $\alpha\Rightarrow_G\beta$, אם $\alpha\in(V\cup T)^*\gamma(V\cup T)^*$ ו-$\beta\in(V\cup T)^*\delta(V\cup T)^*$ כאשר קיים ב-G כלל הגזירה $\delta\to\gamma$. במילים אחרות α,β שייכים ליחס $\Rightarrow_G$ אם ניתן לגזור את β מ-α באמצעות צעד גזירה אחד בלבד.
- באותו אופן נגדיר את היחס $\Rightarrow^i_G$, כך שמתקיים $\alpha\Rightarrow^i_G\beta$ אם ניתן לגזור את β מ-α על ידי i צעדי גזירה בדיוק.
- יחס נוסף $\Rightarrow^*_G$, שבו $\alpha\Rightarrow^*_G\beta$ אם ניתן לגזור את β מ-α על ידי סכום צעדי גזירה כלשהם.

בדקדוק G, אם $\alpha\in(V\cup T)^*$ וגם $S\Rightarrow^*_G\alpha$; נאמר ש-α היא פסוקית ב-G.

ב) ההיררכיה של חומסקי

ההיררכיה של חומסקי מחלקת את כל דקדוקי השכתוב לארבעה דורות $T_3\subseteq T_2\subseteq T_1\subseteq T_0$, שבכל דורה מתווספים עוד אילוצים על הדורה הקודמת, כך שמספר השפות שיכולה לקבל פוחת ומוכל בקבוצת השפות שניתן ליצור מהדורה הקודמת. הסיבה שאנו מתעניינים בחלוקה זו היא שכל קבוצת שפות Γ_i המתאימה לדורה T_i מתאימה למודל חישובי אחר.

T_0 - כל הדקדוקים ללא הגבלות. מקבלת את קבוצת השפות Γ_0 שהיא קבוצת כל השפות שיש עבורן דקדוק כלשהו. באופן מפתיע קבוצה זו מאוד חזקה ומקבילה ל-R_E, שהיא קבוצת כל השפות שיכול להתקבל על ידי מחשב כללי השקול למודל טיורינג, שהוא המודל החישובי הכי חזק שאנו מכירים שניתן לבנות!

T_1 - בדקדוקים אלו מותרים רק כללי גזירה $\alpha\to\beta$ כאשר $|\alpha|\le|\beta|$, כלומר לא ניתן להקטין את המחרוזת האמורה. יוצא דופן הוא כלל הגזירה $S\to\varepsilon$ כדי שנוכל לכלול את ε בשפה*. קבוצת הדקדוקים המתאימים ל-T_1 נקראת "דקדוקים תלויי הקשר", ומקבלת את קבוצת השפות Γ_1 השקולה לקבוצת השפות שניתן לקבל במודל של מכונת טיורינג עם זיכרון לינארי, המשתמש ב-$O(n)$* תאים בסרט. יש טענה שתופסת את כל השפות הטבעיות (אנגלית, עברית, וכו׳).

T_2 - כמו T_1 לא מאפשרת כללים מקצרי אורך, אך בנוסף בכל כלל גזירה $\alpha\to\beta$, α הוא משתנה בודד בלבד. כלומר $\alpha\in V$. קבוצת הדקדוקים המתאימים ל-T_2 נקראים "דקדוקים חסרי הקשר", מכיוון שכלל הגזירה עבור כל משתנה אינו תלוי בסביבתו במחרוזת הנוכחית. T_2 מקבל את קבוצת השפות Γ_2 השקולה לאוטומט מחסנית אי-דטרמיניסטי, עליו נלמד בהמשך.

* נתיר $\varepsilon\to A$ אם $A\in V$. לא אוסיף ... צריך מכוח החישוב.

* n הוא אורך הקלט למכונה.

T_3 - גם כן לא מאפשר כללי גזירה מקדמי אורך ו-α משתנה בודד מ-V, אך הנוסח מאפשר רק כללי גזירה לינאריים ימניים או שמאליים, אך לא שניהם באותו דקדוק. דקדוק לינארי ימני הוא דקדוק שכל כללי הגזירה בו הם מהצורה $\alpha \to aB|a$, והדקדוק לינארי שמאלי $\alpha \to Ba|a$. כאשר $\alpha \in V$, $B \in (V \cup T)^*$, ו- $a \in T$. קבוצת הדקדוקים המתאימים ל-T_3 נקראים "דקדוקים רגולריים", מפני שקבוצת השפות T_3 ש-T_3 מקבלת שקולה לאוטומט סופי דטרמיניסטי המקבל את כל השפות הרגולריות. דקדוק לינארי ימני ושמאלי שקולים כל אחד לחוד לקבוצת השפות הרגולריות.

ה) כיצד מוכיחים נכונות בניה של דקדוק

כאשר נתון דקדוק G ואנו מנחשים שהוא מתאים לשפה L, נצטרך להוכיח הכלה דו-כיוונית בין L ל-$L(G)$. כלומר להוכיח, $w \in L \longleftrightarrow S \Rightarrow^*_G w .\ \forall w \in T^*$. שני הכיוונים יעשו באמצעות אינדוקציה מאחר ולסדרות גזירה יש מבנה רקורסיבי.

$L \subseteq L(G)$: נניח $w \in L$ ונוכיח $S \Rightarrow^*_G w$. ברוב המקרים לא נוכל לעשות זאת באמצעות אינדוקציה ישירה על האורך של w, שכן יכול להיות שכאשר נרצה להשתמש בהנחת האינדוקציה נקבל מילה טרמינלית באורך k ולא נוכל לפתח אותה למילה באורך $k+1$. לכן במקום, נוכיח, גם כן באינדוקציה, טענת עזר האומרת שניתן להגיע מ-S לכל פסוקית α המכילה משתנה, כך ש-α מתאימה ל-w. אם α מוגדר באמצעות איטרציות באורך k יותר נוח שהאינדוקציה תהיה על k, אחרת, תהיה על אורך הפסוקית α. לאחר שנוכיח את טענת העזר, נראה באמצעות מספר צעדי גזירה פשוטים ניתן להגיע מ-α אל w.

כדאי לראות מתי כן ניתן לעשות אינדוקציה על אורך של w.

$$\underbrace{S \Rightarrow^*_G \alpha}_{\text{טענת עזר}} \underbrace{\Rightarrow^*_G w}_{\text{צעדי גזירה}}$$

הערה - בדקדוקים בהם המשתנה S נשאר בכללי הגזירה, כלומר הדקדוקים הם מהצורה $S \to (V \cup T)^* S (V \cup T)^*$, כן נוכל להשתמש באינדוקציה על אורך w, על ידי שניקח את הנחת האינדוקציה ונגדיל אותה בכלל הגזירה, כך להגדיל את הפסוקית עד שנגיע ל-w. אמנם גם במקרים אלו השיטה הכללית תעבוד.

$L(G) \subseteq L$: נניח $w \in L(G)$ ונוכיח $w \in L$. גם כאן נוכיח באינדוקציה טענת עזר שבה כיצד נראות כל הפסוקיות α שאינה מילה טרמינלית, אלא מכילה לפחות משתנה אחד. האינדוקציה תהיה על כמות הגזירות. לאחר הוכחת טענת העזר, ניקח את כל צורות α שמהם ניתן להגיע למילה טרמינלית בצעד גזירה בודד, ונראה שכל המילים הטרמינליות שניתן להגיע אליהן ב-G שייכים גם ל-L.

דוגמה: בהינתן הדקדוק: $G=(V=\{S,A,B\},\ T=\{a,b\},\ S,\ P=\{S \to aB|a,\ A \to aB|a,\ B \to bA|b\})$

נוכיח $L=\{(ab)^k a : k \geq 0\}$

כיוון 1 - יהי $w=(ab)^k a\in L$, צ"ל $S\Rightarrow^*_G w$. עבור $k\ge0$ נקבל $w=a$ ואכן ישנו כלל גזירה $S\to a$. נמשיך בטענת עזר.

טענת עזר: קיימת ב-G סדרת גזירה מהצורה הבאה $S\Rightarrow^*_G (ab)^k A$ לכל $k\ge1$. באינדוקציה על k.

בסיס - עבור $k=1$ נקבל את הפסוקית abA, ואכן יש סדרת גזירה כזו. $S\Rightarrow aB\Rightarrow abA$.

צעד - נניח נכונות הטענה עבור k, ונוכיח עבור k+1. באמצעות סדרת הגזירה הבאה:

$S\underset{\text{הנחת האינדוקציה}}{\Rightarrow^*}(ab)^kA\Rightarrow(ab)^kaB\Rightarrow(ab)^kabA=(ab)^{k+1}A$ מ.ש.ל

באמצעות הטענת עזר נראה שיש סדרת גזירה מ-S ל-w:

$S\Rightarrow(ab)^kA\Rightarrow(ab)^ka$ מ.ש.ל

כיוון 2 - נוכיח שכל מילה טרמינלית ב-G שייכת ל-L. נתחיל בטענת עזר.

טענת עזר: כל פסוקית α שאינה טרמינלית ב-G היא אחת משלושת הצורות הבאות:

$\alpha=S$, $\alpha=(ab)^kA$ עבור $k\ge1$, $\alpha=(ab)^kaB$ כאשר $k\ge0$.

הוכחה: באינדוקציה על מספר הגזירות;

בסיס - עבור $i=0$ צעדי גזירה נקבל $S=\alpha$,

צעד - נניח שהטענה נכונה עבור $0\le i$ ונוכיח עבור i+1. נתבונן על סדרת הגזירה $S\Rightarrow^{i+1}_G\alpha$, ונפרק אותה כך $S\Rightarrow^i_G\beta\Rightarrow\alpha$. לפי הנחת האינדוקציה β היא אחת משלושת הצורות לעיל. נבחן כל צורה עבור β ואת צעדי הגזירה מ-β שעבורם לא נקבל מילה טרמינלית. עבור $\beta=S$ נקבל $\alpha=aB$. עבור $\beta=(ab)^kA$ נקבל $\alpha=(ab)^kaB$, ועבור $\beta=(ab)^kaB$ נקבל $\alpha=(ab)^kabA=(ab)^{k+1}A$. ניתן לראות ש-$\alpha$ עבור כל אחד מהצורות מקיים את הטענה.

מבין שלושת הצורות בטענת עזר, הצורה היחידה שממנה ניתן להגיע למילה טרמינלית בצעד אחד היא $\alpha=(ab)^kA$, וכל מילה טרמינלית שתתקבל היא מהצורה $w=(ab)^ka$. נמצא שכל מילה טרמינלית שניתן להגיע אליה ב-G היא מילה שמתקבלת ב-L. לכן $L(G)\subseteq L$. מ.ש.ל

1) שקילות דקדוקים רגולריים (T3) ו-DFA

הדקדוקים הרגולריים מחולקים לשני סוגים: לינאריים ימניים ולינאריים שמאליים. כיוון ששני הסוגים שקולים, נוכיח שקילות רק בין הדקדוקים הלינאריים הימניים ל-DFA, אך באותו אופן היה ניתן לעשות עבור הדקדוקים הלינאריים השמאליים. גישת ההוכחה תהיה קונסטרוקטיבית, כלומר נראה שמכל DFA ניתן לבנות דקדוק לינארי ימני שקול, וההיפך.

בניית דקדוק מאוטומט

יהי DFA המוגדר $A=(Q,q_0,\Sigma,F,\delta)$. צ"ל קיים דקדוק לינארי ימני G_A כך ש-$L(G_A)=L(A)$. נבנה G_A שתהליך הגזירה בו מחקה את התהליך ש-A מבצע על כל מילה. לשם כך נדרש שיתקיים:

$\forall x\in\Sigma^*$, $\forall p,q\in Q$, $\delta(p,x)=q \iff p\Rightarrow^*_{G_A}xq$, כלומר ב-A לאחר הרצת x מ-p יגיע ל-q, וכן ב-G_A לאחר גזירת x משתנה p יגיע לפסוקית שבה משתנה q בצד ימין. נוכיח זאת בהמשך בטענת עזר.

בס"ד

הבנ'ה של G_A תה'ה מהצורה $G_A=(V=Q, T=\Sigma, S=q_0, P)$ כאשר ל-P יהיו שלושה סוג' כלל' גז'רה:

(1) לכל מעבר $\delta(p,\sigma)=q$ כאשר $p,q\in Q$ ו-$\sigma\in\Sigma$ ב-A, יה'ה ה-G_A את כלל הגז'רה $p\to\sigma q$.

(2) אם $\delta(p,\sigma)=q$ ובנוסף $q\in F$, נוס'ף ה-G_A את כלל הגז'רה $p\to\sigma$. כלל זה יאפשר לדקדוק לעצור את סדרת הגז'רה ולקבל מילה טרמינל'ת כאשר A מג'ע למצב מקבל.

(3) אם $\varepsilon\in L(A)$; נוס'ף ל-G_A את כלל הגז'רה $S\to\varepsilon$ כדי לכלול את ε ב-$L(G_A)$. הוכחה שהוספה זו לא מוס'פה מל'ם מ'ותרות ל-$L(G_A)$ במטלה 2.

טענת עזר 1: יהי G דקדוק ל'נארי ימני, כל הסבירות α של G היא מהצורה $\alpha=wA$ או $\alpha=w$ כאשר $w\in T^*$ ו-$A\in V$. טענה זו נכונה לכל דקדוק ל'נארי ימני. | את עזר 1-2

הוכחה: באינדוקציה על אורך הגז'רה i.

בס'ס- עבור $i=0$ צעד' גז'רה נקבל $\alpha=S$ שהיא אכן מהצורה $\alpha=wA$ עבור $w=\varepsilon$.

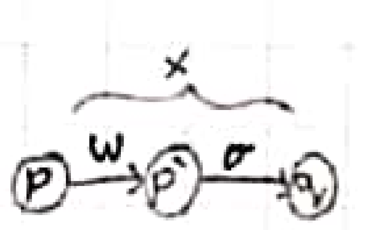
צעד- נניח נכונות הטענה עבור i ונוכיח עבור $i+1$ צעד' גז'רה. נתבונן בסדרת הגז'רה $S\overset{i+1}{\underset{G}{\Rightarrow}}\alpha$ ונפרק אותה ל- $S\overset{i}{\underset{G}{\Rightarrow}}\beta\underset{G}{\Rightarrow}\alpha$. לפי הנחת האינדוקציה $\beta=wA$ (לא יכול להיות $\beta=w$). כיוון ש-G דקדוק ל'נארי ימני, ישנם 3 אפשרויות לכלל גז'רה. באפשרות הראשונה הכלל הוא $A\to\sigma$ ואז $\alpha=w\sigma$. באפשרות השנ'ה $A\to\sigma A$, ואז $\alpha=w\sigma A$. באפשרות השלישית $A\to\varepsilon$ ואז $\alpha=w$. בכל המקרים קיבלנו צורות מהטענה. מ.ש.ל

טענת עזר 2: יהי באוטומט A $p,q\in Q$ ו-$x\in\Sigma^*$ אז' מתק"ים $\delta(p,x)=q \iff p\overset{*}{\underset{G_A}{\Rightarrow}}xq$.

הוכחה: כיוון 1- נניח $\delta(p,x)=q$, נוכ'ח $p\overset{*}{\underset{G_A}{\Rightarrow}}xq$ באינדוקציה על אורך המילה $|x|=i$.

בס'ס- עבור $i=0$ לפי הנחה $\delta(p,\varepsilon)=q$, כלומר $p=q$. ולכן $p\overset{*}{\underset{G_A}{\Rightarrow}}p$ עבור 0 צעד' גז'רה.

צעד- נניח נכונות עבור i ונוכיח עבור אורך המילה $i+1$. כיוון ש-x לא מילה ריקה, נניח כי $x=w\sigma$. כעת נפרק את ההרצה של A על p ו-x לשני חלקים: $\delta(p,x)=\delta(p,w\sigma)=\delta(\delta(p,w),\sigma)=q$.



נסמן $\delta(p,w)=p'$. לפי הנחת האינדוקציה מתק"ים $p\overset{*}{\underset{G_A}{\Rightarrow}}wp'$. לפי הכלל הראשון של כלל' הגז'רה הבנ'ה של G_A יש כלל גז'רה $p'\to\sigma q$. לכן אם נשתמש בכלל זה נקבל שמתק"ים: $p\overset{*}{\underset{G_A}{\Rightarrow}}wp'\underset{G_A}{\Rightarrow}w\sigma q=xq$, כנדרש. | זהו ישנו המעבר $\delta(p',\sigma)$, אחרת לא יכלנו להג'ע ל-q.

כיוון 2- נניח $p\overset{*}{\underset{G_A}{\Rightarrow}}xq$ ונוכ'ח $\delta(p,x)=q$. נוכ'ח באינדוקציה על אורך הגז'רה.

בס'ס- עבור $i=0$, לפי הנחה $p\overset{0}{\underset{G_A}{\Rightarrow}}xq$. בהכרח $p=q$ ו-$x=\varepsilon$. אכן מתק"ים $\delta(p,\varepsilon)=q$.

צעד- נניח שהטענה נכונה עבור i ונוכיח עבור $i+1$. לפי הנחה מתק"ים $p\overset{i+1}{\underset{G_A}{\Rightarrow}}xq$. נפרק את סדרת הגז'רה ל- $p\overset{i}{\underset{G_A}{\Rightarrow}}\beta\underset{G_A}{\Rightarrow}xq$. מטענת עזר 1 בהכרח $\beta=wp'$ כי א'נה טרמינל'ת. מהנחת האינדוקציה עבור β ק"ים המעבר $\delta(p,w)=p'$. בנוסף כדי להג'ע בסדרת הגז'רות מ-β ל-α ח'יב להיות מוגדר כלל הגז'רה $p'\to\sigma q$ כך ש-$x=w\sigma$. מהגדרת הבנ'ה של G_A נוכל להס'ק כי ק"ים המעבר $\delta(p',\sigma)=q$. בפירוט כל המעברים ב-A ק"ים השרשור הבא:

$\delta(p,x)=\delta(p,w\sigma)=\delta(\delta(p,w),\sigma)=\delta(p',\sigma)=q$ מ.ש.ל

נחזור להוכחה הכללית, נוכיח $L(G_A)=L(A)$ בהכלה דו-כיוונית.

כיוון 1 - יהי $x\in L(A)$, נראה $x\in L(G_A)$. אם $x=\varepsilon$ אזי בבניית G_A יהיה כלל הגזירה $S\to\varepsilon$, וסיימנו. אחרת נניח $x\neq\varepsilon$. לפי הגדרת DFA מתקיים $\delta(q_0,x)=p$, כאשר $p\in F$. נניח כי $x=w\sigma$, כיוון שהוא שונה מהמילה הריקה, ונפרק את הריצה של A על x, $\delta(q_0,x)=\delta(q_0,w\sigma)=\delta(\delta(q_0,w),\sigma)=p$. נסמן $\delta(q_0,w)=p'$. מטענת עזר 2 ב-G_A קיימת סדרת הגזירה $S \overset{*}{\underset{G_A}{\Rightarrow}} wp'$. כעת, כדי להגיע מ-$p'$ אל p נוכל להסיק כי גם קיים המעבר $\delta(p',\sigma)=p$, ואכן כיוון ש-$p\in F$, לפי ההגדרה השנייה של כללי גזירה ב-G_A, קיים גם הכלל $p'\to\sigma$. כעת נוכל לקבל את סדרת הגזירה $S \overset{*}{\underset{G_A}{\Rightarrow}} wp' \Rightarrow w\sigma = x$. נמצא כי אכן $x\in L(G_A)$.

כיוון 2 - יהי $x\in L(G_A)$, נראה $x\in L(A)$. אם $x=\varepsilon$ אזי בהכרח $x\in L(A)$. אחרת נניח $x\neq\varepsilon$. מההנחה $x\in L(G_A)$ נובע כי ישנה סדרת גזירות $S \overset{*}{\underset{G_A}{\Rightarrow}} x$. נניח כי $x=w\sigma$, ונפרק את סדרת הגזירה $S \overset{*}{\underset{G_A}{\Rightarrow}} B \underset{G_A}{\Rightarrow} x$. לפי טענת עזר 1 נוכל להניח כי $B=wp$. ומטענת עזר 2 נוכל להסיק כי ב-A קיים המעבר $\delta(q_0,w)=p$. בנוסף כיוון ש-x מילה טרמינלית חייב להיות ב-G_A כלל הגזירה $p\to\sigma$, ולכן גם חייב להיות המעבר $\delta(p,\sigma)=q$ כאשר $q\in F$, אחרת $p\to\sigma$ לא היה נכנס ל-G_A לפי הגדרת הבנייה. כעת נוכל לקבל את סדרת המעברים

$$\delta(q_0,x) = \delta(q_0,w\sigma) = \delta(\delta(q_0,w),\sigma) = \delta(p,\sigma) = q\in F \Rightarrow x\in L(A)$$ מ.ש.ל.

בניית אוטומט מדקדוק

יהי דקדוק לינארי ימני $G=(V,T,S,P)$, נראה קיים DFA A_G כך ש-$L(A_G)=L(A)$. גם כאן ההוכחה תהיה קונסטרוקטיבית, אלא שנבנה ε-NFA ולא DFA. אמנם כיוון שהם שקולים אין הבדל בע"ה. נבנה A_G שתהליך החישוב בו מחקה את תהליך הגזירה של G עם כל פסוקית ב-G. לשם כך נרצה שיתקיים:

$A\in\delta(q_0,x) \Leftrightarrow S \overset{*}{\underset{G}{\Rightarrow}} xA \quad \forall A\in V, \forall x\in\Sigma^*\setminus\{\varepsilon\}$. כלומר ניתן לגזור את xA ב-G אם"ם ב-A_G נוכל להגיע למצב A ע"י חישוב של x. נוכיח בטענת עזר.

הבנייה של A_G תהיה מהצורה $A_G=(Q=V\cup\{q_f\}, q_0=S, \Sigma=T, F=\{q_f\}, \delta)$, כאשר q_f הוא מצב יחיד ב-Q המקבל מילה ב-A_G. את פונקציית המעברים δ נגדיר לפי ארבעת השלבים הבאים:

(1) לכל $q\in Q$ ו-$\sigma\in\Sigma\cup\{\varepsilon\}$ נאתחל כל מעבר ל-$\delta(q,\sigma)=\emptyset$.

(2) לכל כלל גזירה ב-G מהצורה $A\to aB$ כאשר $A,B\in V$ ו-$a\in T$, נוסיף ל-$\delta(A,a)$ את B.

(3) לכל כלל גזירה ב-G מהצורה $A\to a$ כאשר $A\in V$ ו-$a\in T$, נוסיף ל-$\delta(A,a)$ את q_f.

(4) אם $\varepsilon\in L(G)$, כלומר יש כלל גזירה $S\to\varepsilon$ ב-G, נוסיף ל-$\delta(q_0,\varepsilon)$ את q_f.

<u>טענת עזר 3:</u> יהי $x\in\Sigma^*\setminus\{\varepsilon\}$ ו-$A\in V$. מתקיים $S \overset{*}{\underset{G}{\Rightarrow}} xA$ אם ורק אם $A\in\delta(q_0,x)$.

הוכחה: כיוון 1 - נניח $A\in\delta(q_0,x)$ ונוכיח $S \overset{*}{\underset{G}{\Rightarrow}} xA$ באינדוקציה על $|x|=i$. כיוון ש-$x\neq\varepsilon$ נתחיל מ-$i=1$.

בסיס - עבור $i=1$ נקבל $x\in\Sigma$. לכן מהשלב השני בבניית A_G, כיוון שקיים המעבר $A\in\delta(q_0,x)$, אזי ב-G בהכרח קיים כלל הגזירה $S\to xA$.

צעד - נניח שהטענה נכונה עבור i ונוכיח עבור $i+1$. נסמן $x=w\sigma$, ונפרק את מסלול החישוב של A_G על x כך: $\delta(q_0,x)=\delta(q_0,w\sigma)=\delta(\delta(q_0,w),\sigma)=A$. נסמן $\delta(q_0,w)=B$. כיוון ש-$|w|=i$, לפי הנחת האינדוקציה מתקיים $S\overset{*}{\Rightarrow}_G wB$. כיוון ש-$A\in\delta(B,\sigma)$ גם קיים ב-G כלל הגזירה $B\to\sigma A$. כעת נוכל לקבל את סדרת הגזירות הבאה: $S\overset{*}{\Rightarrow}_G wB\Rightarrow_G w\sigma A=xA$.

כיוון 2 - הוכחתי בשאלה 2.

<u>טענת עזר 4:</u> יהי $x\neq\varepsilon$. מתקיים $S\overset{*}{\Rightarrow}_G x$ אם ורק אם $q_F\in\delta(q_0,x)$. כלומר $x\in L(G)$ אם"ם $x\in L(A_G)$.

הוכחה: כיוון 1 - יהי $x\neq\varepsilon$ כך שמתקיים $S\overset{*}{\Rightarrow}_G x$, נראה $q_F\in\delta(q_0,x)$. נפרק את סדרת הגזירה $S\overset{*}{\Rightarrow}_G B\Rightarrow_G x$. מטענת עזר 1 B הוא מהצורה $B=wA$. נחלק לשני מקרים:

(1) אם $w\neq\varepsilon$, אזי כיוון שמתקיים $S\overset{*}{\Rightarrow}_G wA$ מטענת עזר 3 גם מתקיים $A\in\delta(q_0,w)$. כעת אם $A=S$ והשתמשנו בכלל הגזירה $S\to\varepsilon$ כדי לקבל מילה טרמינלית, כלומר $w=x$, אזי לפי שלב 4 בבנייה של A_G, $q_F\in\delta(q_0,\varepsilon)$. ומכאן נוכל לקבל את מסלול החישוב הבא: $\delta(q_0,x)=\delta(q_0,w\varepsilon)=\delta(\delta(q_0,w),\varepsilon)=\delta(q_0,\varepsilon)=q_F$

אחרת אם $A\neq S$, אזי היינו צריכים להשתמש בצעד גזירה האחרון בכלל $A\to\sigma$, כאשר $x=w\sigma$. משלב 3 בבניית A_G מתקיים $q_F\in\delta(A,\sigma)$. כעת נבנה את מסלול החישוב:

$$\delta(q_0,x)=\delta(q_0,w\sigma)=\delta(\delta(q_0,w),\sigma)=\delta(A,\sigma)=q_F$$

(2) $w=\varepsilon$, ולכן בהכרח $B=S$, כלומר הגענו אל B ב-0 צעדי גזירה. המסקנה היא שסדרת הגזירה מ-S ל-x היא צעד גזירה בודד, $S\Rightarrow_G x$, ולכן חייבים להשתמש בכלל הגזירה $S\to\sigma$ כך ש-$x=\sigma$. משלב 3 של הגדרת הבנייה $q_F\in\delta(q_0,\sigma)$. מסלול החישוב הינו $\delta(q_0,x)=\delta(q_0,\sigma)=q_F$

כיוון 2 - יהי $x\neq\varepsilon$ כך שמתקיים $q_0\in\delta(q_0,x)$, נראה $S\overset{*}{\Rightarrow}_G x$. נניח כי $x=w\sigma$, ונפרק את המסלול המקבל: $\delta(q_0,x)=\delta(q_0,w\sigma)=\delta(\delta(q_0,w),\sigma)=q_F$. נסמן $\delta(q_0,w)=A$. באופן דומה לכיוון הראשון, ייתכנו שני מקרים: (5)

(1) אם $w\neq\varepsilon$. אזי מטענת עזר 3, בגלל שמתקיים $A\in\delta(q_0,w)$, גם מתקיים $S\overset{*}{\Rightarrow}_G wA$. כעת, אם הצעד האחרון קורא $\sigma=\varepsilon$, אזי בהכרח $A=S$ ו-$x=w$. ולכן ב-G קיים כלל הגזירה $S\to\varepsilon$. סדרת הגזירות הרצויה ב-G היא אם כן: $S\overset{*}{\Rightarrow}_G wS\Rightarrow_G w=x$

אחרת אם $\sigma\neq\varepsilon$, אזי כדי ש-x תתקבל בהכרח $q_F\in\delta(A,\sigma)$, ומשלב 3 של הגדרת הבנייה, בהכרח ב-G קיים הכלל $A\to\sigma$. סדרת הגזירה הרצויה היא: $S\overset{*}{\Rightarrow}_G wA\Rightarrow w\sigma=x$

(2) אם $w=\varepsilon$, אזי $\sigma\neq\varepsilon$ שהרי הגדרנו $x\neq\varepsilon$, ולכן $x=\sigma$. כדי ש-x תתקבל ב-A_G צריך שיתקיים $q_F\in\delta(q_0,\sigma)$. משלב 3 של הבנייה, קיים ב-G הכלל $S\to\sigma$. נוכל כעת לקבל את סדרת החישוב הרצויה: $S\Rightarrow_G\sigma=x$

טענת עזר 5: מתקיים $S \Rightarrow^*_G \varepsilon$ אם ורק אם $q_F \in \delta(q_0,\varepsilon)$. כלומר $\varepsilon \in L(G)$ אם ורק אם $\varepsilon \in L(A_G)$.

הוכחה: כיוון 1 - נניח $S \Rightarrow^*_G \varepsilon$. הדרך היחידה להגיע מ-S ל-ε ב-G בדקדוק לינארי ימני, הוא כאשר קיים כלל הגזירה $S \to \varepsilon$. לכן אז לפי שלב 4 של בניית A_G $q_F \in \delta(q_0,\varepsilon)$.

כיוון 2 - נניח $q_F \in \delta(q_0,\varepsilon)$. מהגדרת הבנייה של A_G זה יכול לקרות רק כאשר ב-G קיים $S \to \varepsilon$, נמצא כי אכן ב-G ניתן לגזור $S \Rightarrow^*_G \varepsilon$.

נחזור להוכחה הכללית. ניתן לראות כי משילוב טענות עזר 4 ו-5, לכל $x \in \Sigma^*$ מתקיים $x \in L(G)$ אם ורק אם $x \in L(A_G)$.

ג) עצי גזירה

יהי דקדוק חסר הקשר $G=(V,T,S,P)$. נאמר ש-T_G הוא עץ גזירה עבור G אם הוא עץ מכוון וסדור בעל שורש המקיים את חמשת הכללים הבאים:

(1) כל צומת ב-T_G מסומנת בסימן הלקוח מתוך $V \cup T \cup \{\varepsilon\}$.

(2) השורש יסומן ב-S. לפעמים נרצה לבחון תת-עץ של משתנה כלשהו ואז נסכים שהשורש יהיה משתנה זה.

(3) כל צומת פנימית שאינה עלה תסומן במשתנה מ-V.

(4) אם קיימת צומת פנימית המסומנת במשתנה A ובניה הם $X_1,X_2,\dots,X_n$. אזי ב-G קיים כלל הגזירה $A \to X_1X_2\dots X_n$

(5) אם קיים קודקוד המסומן ב-ε, אזי קודקוד זה הינו עלה שהוא הבן היחיד לאביו.

נגדיר "חזית של עץ גזירה" להיות שרשור הסימנים של כל העלים ע"י משמאל לימין. החזית אינה חייבת להיות טרמינלית.

טענה: קיים עץ גזירה עבור דקדוק G עם חזית α אם"ם $S \Rightarrow^*_G \alpha$. כלומר כל עץ גזירה של G החזית α היא פסוקית של G

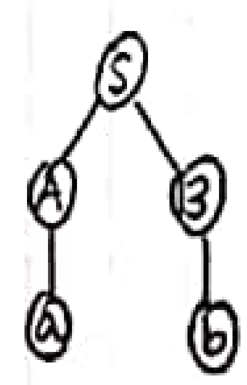
אם תת עץ של עץ גזירה של G אזי הוא עץ גזירה שמייצג את כל הפסוקיות שניתן להגיע אליהן מהמשתנה בשורש של התת-עץ. הטענה הנ"ל נכונה גם עבור תת-עץ גזירה. קיים עץ גזירה עבור G ששורשו A וחזיתו α אם"ם $A \Rightarrow^*_G \alpha$.

השימוש העיקרי בעצי גזירה הוא כאשר יש לפסוקית אחת כמה סדרות גזירה הנבדלות רק בסדר הפעלת כללי הגזירה, אנו רוצים לראות אותן כגזירות שקולות. ואכן עץ הגזירה עבור כל סדרת גזירה כזו יהיה זהה.

דוגמא: עבור הדקדוק G שכללי הגזירה בו הן: $P=\{S \to AB,\ A \to a,\ B \to b\}$ נסתכל על

סדרות הגזירה: $S \Rightarrow AB \Rightarrow aB \Rightarrow ab$ שמתקבלת בהן אותה פסוקית ואך סדר גזירה שונה.
$S \Rightarrow AB \Rightarrow Ab \Rightarrow ab$

עץ הגזירה עבורן יהיה זהה.



(18)

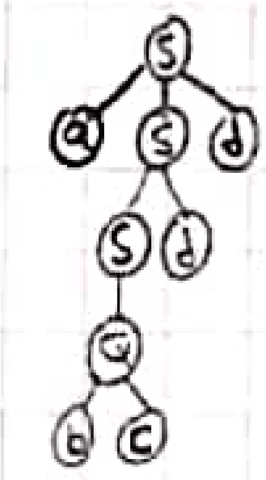
ח) דו - משמעות

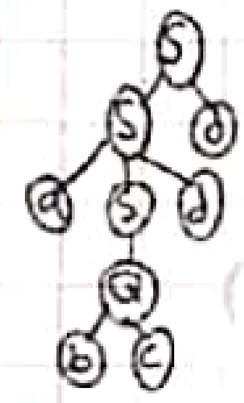
<u>דקדוק דו - משמעי</u>: דקדוק חסר הקשר G יקרא "דו-משמעי" אם ניתן לגזור ממנו שני עצי גזירה שונים עבור תבנית כלשהי. כלומר יש ב-G פסוקית α שניתן לקבל אותה בשתי סדרות גזירה שונות ושאינן נבדלות רק בסדר הפעלת כללי הגזירה.

דוגמה: הדקדוק G שכללי הגזירה בו הן: $P=\{S \to aSd \mid Sd \mid Q,\ Q \to bQc \mid bc\}$, ומקבל את השפה $L=\{a^l b^m c^m d^n : m \geq 1, 0 \leq l \leq n\}$, הינו דו-משמעי שכן את המילה abcdd ניתן לקבל בשתי סדרות גזירה שונות עם עצי גזירה שונים:

(1) $S \Rightarrow aSd \Rightarrow aSdd \Rightarrow aQdd \Rightarrow abcdd$

(2) $S \Rightarrow Sd \Rightarrow aSdd \Rightarrow aQdd \Rightarrow abcdd$.

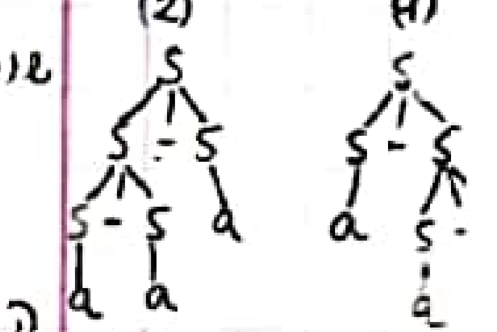
(1) 

(2) 

בבניית דקדוק לשפה, נרצה שהדקדוק יהיה חד-משמעי ולא דו-משמעי, הסיבה לכך היא שבשפות מסוימות יכול להיות לסדר גזירת המילה משמעות חשובה, כך שאם יהיו מספר סדרות כאלו נקבל תוצאה שגויה. לדוגמה, קומפיילר שמנסה לפענח ביטוי אריתמטי $a-a-a$, שנבנה באמצעות הדקדוק $S \to S-S \mid a$, נקבל תוצאות שונות בהתאם לסדרת הגזירה בה השתמשנו:

(1) $S \Rightarrow S-S \Rightarrow a-S \Rightarrow a-S-S \Rightarrow a-a-S \Rightarrow a-a-a = a-(a-a)=a$

(2) $S \Rightarrow S-S \Rightarrow S-S-S \Rightarrow a-S-S \Rightarrow a-a-S \Rightarrow a-a-a = (a-a)-a=-a$

(2) (1)


<u>הפיכת דקדוק דו-משמעי לחד-משמעי</u>: לא כל דקדוק דו-משמעי ניתן להפוך לחד-משמעי, אמנם את אלה שכן, הדו-משמעות נובעת מכך שאין סדר על "יצור האיברים" של הפסוקית. הקו המנחה הוא לפרק כל שלב בגזירת הפסוקית, כך שלכל שלב אחראי משתנה אחר. כאשר משתנה אחד מסיים את תפקידו הוא מעביר שליטה למשתנה אחר האחראי על השלב הבא.

לדוגמה: עבור הדקדוק הדו-משמעי שהראינו לפני, המתאים לשפה $L=\{a^l b^m c^m d^n : m \geq 1, 0 \leq l \leq n\}$, נגדיר משתנה חדש B, כך ש-S יוסיף a,d בצורה שווה, ואז יעביר שליטה ל-B, שיוסיף כמות בלתי מוגבלת של d ויעביר שליטה ל-Q שיוסיף כמות שווה של b,c. $P=\{S \to aSd \mid B,\ B \to Bd \mid Q,\ Q \to bQc \mid bc\}$.

<u>שפה דו-משמעית</u>: שפה חסרת הקשר תיקרא "דו-משמעית" אם כל דקדוק היוצר אותה הוא דו-משמעי. לא נלמד להוכיח האם שפה היא דו-משמעית.

להוכיח שדקדוק הוא ... משמעי, יש להוכיח ... את ... כי ... נראה ... מילה טרמינלית בשפה ... זוגית. ... אבן להוכיח שלכל ... השפה קיימת סדרת ... גזירה שמאלית בודדה יחידה. ... זה תהיה באינדוקציה ... אורך המילה, ותדרוש ... כל צורה של מילה ... עם ... מתאימות ... דוגמה ... לא מלאה ... לא עלינו.

ט) הצורה הנורמלית של חומסקי

לכל שפה חסרת הקשר L קיים דקדוק חסר הקשר לשפה $L'=L\setminus\{\varepsilon\}$ כך שהדקדוק הוא מהצורה הנורמלית של חומסקי. צורה זו מתירה רק שני סוגים של כללי גזירה.

(1) $A \to a$ כך ש-$a \in T$ ו-$A \in V$.

(2) $A \to BC$ כך ש-$A,B,C \in V$.

דוגמה: נבנה דקדוק חסר הקשר מהצורה הנורמלית של חומסקי לשפה $L=\{a^n b^n : n \in \mathbb{N}\}$. הדקדוק הוא:
$G=(V=\{S,X,A,B\},\ T=\{a,b\},\ S,\ P=\{S \to AB \mid AX,\ X \to SB,\ A \to a,\ B \to b\})$

אלגוריתם להעברת דקח"ה לצורה הנורמלית של חומסקי:

(1) נסלק כללי גזירה המכילים ε.

(2) נסלק כללים מהצורה $A \to B$ כאשר $A,B\in V$.

(3) נוסיף כללי גזירה במידת הצורך, כך שלכל $\sigma\in T$ יהיה כלל גזירה $A\to\sigma$.

(4) עבור כל כלל מהצורה $A\to X_1X_2\dots X_n$, אם ישנו $X_i\in T$ נחליף אותו במשתנה המתאים לו משלב (3).

(5) עבור כל כלל מהצורה $A\to X_1X_2\dots X_n$ כאשר $n>2$ נחליף אותו בסדרת הכללים: $A\to X_1Y_1,\ Y_1\to X_2Y_2,\dots,Y_{n-2}\to X_{n-1}X_n$.

4) למת הניפוח לשפות חסרות הקשר

תהי L שפה חסרת הקשר, אזי קיים $n\in \mathbb{N}^+$ כך שלכל $z\in L$ $n\le|z|$, קיים פירוק $z=uvwxy$ המקיים:

(1) $|vwx|\le n$

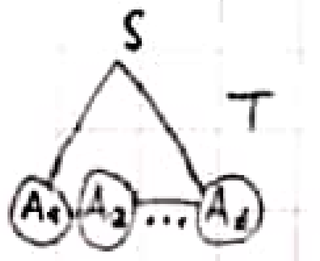
(2) $|vx|\ge 1$

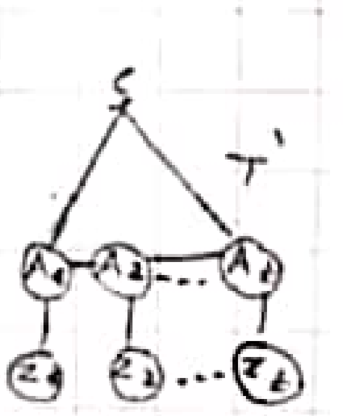
(3) לכל $i\ge 0$ מתקיים: $z^{(i)}=uv^iwx^iy\in L$. v^i ו-x^i הן המילים המתקבלות לאחר שרשור v ו-x i פעמים.

הערות: (1) משפט זה אינו אם ורק אם, לכן לא ניתן להוכיח ממנו ששפה היא חסרת הקשר, שכן ישנם שפות שאינן חסרות הקשר ומקיימות את למת הניפוח. השימוש יהיה כאשר נרצה להוכיח ששפה אינה חסרת הקשר, ואזי נוכל לנסות למצוא מילה שמוכיחה שאינה מקיימת את למת הניפוח. סגנון ההוכחה יהיה בסגנון של הנחה בשלילה, כלומר נניח שהשפה חסרת הקשר ואז נמצא z בשפה שאינו מקיים את למת הניפוח, וזוהי סתירה!

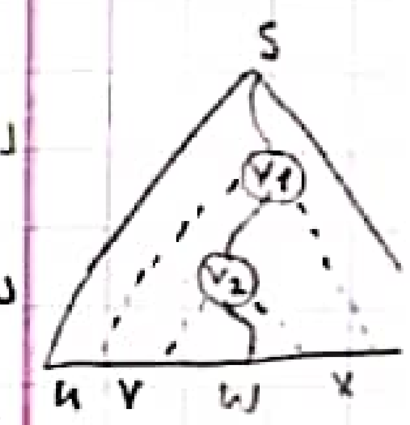
(2) בשפות חסרות הקשר סופיות ישנה לכאורה בעיה, שהרי תנאי 3 הלמה מתקיים עבור אינסוף מילים בשפה? אולם זוהי אינה בעיה שכן נוכל לבחור n שגדול מהמילה הכי ארוכה ב-L. בהכרח יש n כזה שכן השפה סופית. בעת כיוון שאין ב-L מילים באורך זה, למת הניפוח מתקיימת באופן ריק.

<u>הוכחה:</u> תהי שפה חסרת הקשר L, ויהי G דקח"ה בצורה הנורמלית של חומסקי עבור $L\setminus\{\varepsilon\}$. נמצא n כך שלכל $z\in L$ $n\le|z|$ יתקיימו כל שלושת התנאים.





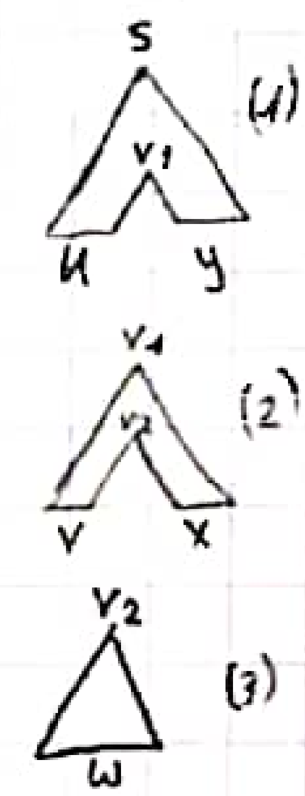
יהי k מספר המשתנים ב-G. נבחר $n=2^k$. תהי $z\in L$ מילה באורך לפחות n. נסמן ב-T את עץ הגזירה של z. ב-T יש t עלים $z_1,z_2,\dots,z_t$ כאשר $n\le t$. כל העלים הינם בנים יחידים של $A_1,A_2,\dots,A_t$ בהתאמה. נבנה עץ T' הזהה ל-T רק שהורדנו ממנו כל העלים. כיוון שהדקדוק הוא מהצורה הנורמלית של חומסקי, T' הוא עץ בינארי, ולכן גובהו h מקיים: $h\ge\log(|z|)\ge\log(2^k)=k$. ומכאן שיש בעץ מסלול באורך h. נתבונן במסלול זה המתחיל מאחד העלים A_i ומטפס כלפי מעלה ועוצר כאשר פוגש צומת המייצגת משתנה שכבר ראה. בהכרח יש צומת כזו שכן המסלול לפחות באורך k ואזי יש בו לפחות $k+1$ צמתים. כיוון שיש רק k משתנים, לפי עקרון שובך היונים, בהכרח יתקבל כפילות באותה צומת.



נסמן ב-v_1 את הצומת השנייה שמצאנו וב-v_2 את הצומת הראשונה שמצאנו. בה"כ שתיהן מסומנות ב-$A\in V$.
נסמן T_1' התת-עץ שמתחיל ב-v_1 ו-T_2' התת עץ שמתחיל ב-v_2. נפרק את z לפי T_1' ו-T_2'. כל החזית של T שמשמאל ל-T_1' היא u, ומימין y. החזית שנותרה שמשמאל ל-T_2' היא v ומימין x. החזית ב-T_2' עצמה היא w.
נשים לב כי מאחר והצירוף הוא של כל החזית, אזי $z=uvwxy$. נוכיח שמתקיימים שלושת התנאים.

(1) צ"ל $|vwx|\le n$. vwx זוהי החזית של T_1'. גובהו של תת-עץ זה הוא לכל היותר k, שהרי לאחר k צעדים בהכרח נבוא למשתנה שכבר ראינו. כיוון שהעץ בינארי, אורך החזית שלו לכל היותר $2^k=n$ כנדרש.

(2) צ"ל $1\le|vx|$. מספיק להוכיח $v\ne\varepsilon$ או $x\ne\varepsilon$. נשים לב: כיוון ש-G מהצורה הנורמלית של חומסקי חייבים להיות לו שני בנים. באחד משני תתי-העצים המתחילים משני בנים אלו מופיע v_2. לכן לתת-עץ שמתחיל מהבן השני החזית שלו אינה ריקה, ומכאן ש-v או x אינם ריקים, כלומר $x\ne\varepsilon$ או $v\ne\varepsilon$.



(3) צ"ל לכל $i\ge 0$ $z^{(i)}=uv^iwx^iy\in L$. נסתכל על שלושה עצי גזירה ב-G. הראשון להם עץ הגזירה של z אלא שבחרנו לגזור את v_1 כעלעלים, כלומר קיצצנו מ-T את T_1'. השני עץ גזירה ששורשו v_1 אלא שבחרנו לגזור את v_2 כעלעלים, כלומר קיצצנו מ-T_1' את T_2'. השלישי הוא עץ הגזירה ששורשו v_2, כלומר T_2'.
משלושת עצי גזירה אלו, נובעים שלושת סדרות הגזירה בהתאמה: $\underset{(1)}{S\Rightarrow^*_G uAy}$, $\underset{(2)}{A\Rightarrow^*_G vAx}$, $\underset{(3)}{A\Rightarrow^*_G w}$.

טענת עזר: לכל $i\ge 0$ מתקיים $A\Rightarrow^*_G v^iAx^i$.

הוכחה: באינדוקציה על i

בסיס - עבור $i=0$ צ"ל $A\Rightarrow v^0Ax^0=A$ ואכן מתקיים בכל דקדוק שהוא.

צעד - נניח נכונות עבור i ונוכיח עבור $i+1$. אכן מתקיים: $A\underset{\text{הנחת האינדוקציה}}{\Rightarrow^*_G} v^iAx^i\underset{(2)}{\Rightarrow} v^ivAxx^i=v^{i+1}Ax^{i+1}$

נחזור להוכחה. יהי $i\in\mathbb{N}$, מתקיים: $S\underset{(1)}{\Rightarrow} uAy\underset{\text{טענת עזר}}{\Rightarrow} uv^iAx^iy\underset{(3)}{\Rightarrow} uv^iwx^iy=\underset{\text{מ.ש.ל}}{z^{(i)}}$