

'Measuring Similarities and Visualising Patterns in a Text Tradition using Pairwise Sequence Alignment and t-Distributed Stochastic Neighbour Embedding'

Estara J Arrant

Jeroen Verrijssen

Appendix: Misalignment Typology

Figure 1: Mathematically Insignificant Misalignment due to Small Word or Letter Insertions

<DirEntry 't1610i_Ex15.16.txt'> <DirEntry 't1652i_Ex15.16.txt'>

תפ-לעליהו נא ימתאמות ואבראבתק וקידרעה גברו-תז-אייתתקו נא
תפ-ולעליהו נא ימתאמות ואבראבתק וקידרעה גברו-תז-אייתתקו נא
תפ-ולעליהו נא ימתאמות ואבראבתק וקידרעה גברו-תז-אייתתקו נא

בנ-י-א: עדzmanד יערו עמק ייינית החליארנו נא-עדzmanד יערו עמק
בנ-י-א: עדzmanד יערו עמק ייינית החליארנו נא-עדzmanד יערו עמק

מזה אילן זדרני יתא ימתגוי זת היד' גבקא
מזה אילן זת היד' גבקא-י-ת-מזה אילן זת היד' גבקא

90

0.6923076923076923

Here, an alignment between t1610i and t1652i at Ex. 15:16, the misalignments occur at ות and at ית because of inserted words. If they were corrected, the score would be unchanged because the number of matches to mismatches would be unchanged.

Figure 2: A Mathematically Significant Alignment

<DirEntry 't1629a_Ex15.15.txt'> <DirEntry 't1681a_Ex15.15.txt'>

בכין אייתבה היילו רברבן יאדום מאיתק יפימואבא-יאחדתנו נו רתיתא:
בכ-�-תבה היילו רברבן יאדום אימק יפ-�-אחדת תי---תנה גונרטיתא:
א

יתברוליבהו נובג-ו פיהו נ-כלדהו נ-יתבי נ-דייריא רענדכנע:
תברוליבהו נובג-ו פיהו נ-כלדהו נ-יתבי נ-דייריא רענדכנע:
תברוליבהו נובג-ו פיהו נ-כלדהו נ-יתבי נ-דייריא רענדכנע:

45

0.4639175257731959

Here, the misalignment occurs in t1681a's word, which has been aligned (with false mismatches) to מואבא and not to the correct correspondent, אחדתנו. These false mismatches artificially lower the alignment score.

Figures 3 and 4: Passage Insertions

ת---א---חֲדָהוֹת---א---מ

רַבִּינוּ: וְדָכְהוּ וּמְרַבֵּי יְהִידָּרְןְלְבָתְיְהּוּן: וְכָדָחַז יְהִי אִידָּאַתְקִיפָּבָב
ר---ה

יְנִיגְלָלְיוּ יְאֵהוּן יְדָעָנוֹנְיִתְהָה: וְפַתְחַזְׁכּוֹלְהָוּנְפּוֹמְהָוּנְכְּחָדָאָוּמְרִין:

תּוֹקְפָּנוּ וּרְבָתּוּשְׁבָחָנְדָחַזְׁיְלְכָלְעָלְמִיְאָכְלָלְיִהְהָאָרְבָּעָכִיתְיְנָאָיְתְּעָבִידְוּבָבָל

נְיִשְׂרָאֵלְכְּדָהְוּוּקְיְמִינְעָלְיִמְאָדְסָוָף: כִּיְתָאָחָדָהוּתְאָמְרָהָנְחָזְזָרְלָמֶץ
א---מ---חֲדָהוֹתְאָמְרָהָנְחָזְזָרְלָמֶץ

רִים: וְכִיְתָאָחָדָהוּתְאָמְרָאַנְיְפּוֹלְלִימָא: וְכִיְתָאָחָדָהוּתְאָמְרָהָנְסְדָרְלָקְה
רִים: 1---חֲדָהוֹתְאָמְרָהָנְסְדָרְלָק

ת---א---חֲדָהוֹת---א---מ

רַבִּינוּ: וְדָכְהוּ וּמְרַבֵּי יְהִידָּרְןְלְבָתְיְהּוּן: וְכָדָחַז יְהִי אִידָּאַתְקִיפָּבָב
ר---ה

יְנִיגְלָלְיוּ יְאֵהוּן יְדָעָנוֹנְיִתְהָה: וְפַתְחַזְׁכּוֹלְהָוּנְפּוֹמְהָוּנְכְּחָדָאָוּמְרִין:

תּוֹקְפָּנוּ וּרְבָתּוּשְׁבָחָנְדָחַזְׁיְלְכָלְעָלְמִיְאָכְלָלְיִהְהָאָרְבָּעָכִיתְיְנָאָיְתְּעָבִידְוּבָבָל

נְיִשְׂרָאֵלְכְּדָהְוּוּקְיְמִינְעָלְיִמְאָדְסָוָף: כִּיְתָאָחָדָהוּתְאָמְרָהָנְחָזְזָרְלָמֶץ
א---מ---חֲדָהוֹתְאָמְרָהָנְחָזְזָרְלָמֶץ

רִים: וְכִיְתָאָחָדָהוּתְאָמְרָאַנְיְפּוֹלְלִימָא: וְכִיְתָאָחָדָהוּתְאָמְרָהָנְסְדָרְלָקְה
רִים: 1---חֲדָהוֹתְאָמְרָהָנְסְדָרְלָק

Between t1610i's *Arba' Kitin* passage occurs in a slightly different location than t1647i's, due to shorter *tosefta* material. The algorithm could not find the correct portion of the text in t1647i, which occurs at the end of the second to last line here.

Figure 5: Gapping to Create Matches

שרה לית מצראים מיתין רמאין על כוֹף ימָא: וְחַמּוֹן יִשְׁרָאֵל יִתְגַּבּוּרָתִיךְ
 שָׁרָאֵל יִתְמַצֵּר אֶימְיִיתִין---עַל כוֹף ימָא:

את קי פָּאָד עֲבָדָה הַיִּינְיִסְיִוּ בְּמַצְרִים: וְדָחַי לְעַמָּמְקָדְמִי יִי וְהַאמְיִנְיִשְׁ
 בְּשֻׁסְמִי מְרָאֵד יִי וּבְנְבִי אֹתִי הַדְּמָשָׁה הַעֲבָדִיָּה: האַבְכָּן שְׁבָחָמְשָׁה וּבְנִישְׁ
 דְּמָשָׁה וּבְנִישְׁ

t1685a's does not occur in t1652 (top sequence). The algorithm has sought to find matches, incorrectly, for this word.

Figure 6: A Noisy yet Valid Alignment

שרה לית שְׁבָחָשׁ יִרְתָּא הַדְּ—אַקְדָּמִי יִי וְאַמְרִי נְלִימִי מְרָנָנוּ דָהּוּ נְשָׁבָחָקְדָמִי יִי
 שָׁרָאֵל יִתְשָׁבָחָשׁ יִרְתָּא—דָהּאַקְדָּמִי יִידְאַמְרוֹ—לִימִי מְרָנָנוּ דָהּוּ נְשָׁבָחָקְדָמִי יִי
 רְמָאַד מְתַגָּאֵי—עַל בַּיִוְתָן יְאוּמָת נְטָל—עַל מְנְטָל—יא: וְכָל מְ—נְדָמָת גָּאָה
 קְדָמִי הַהְוָא בְּמִירִי הַמְתִפְרָעָמִי נִי הַסּוּסְוָן וְרוּכְבִּיהַוּנְעַלְדָא—תַּגָּאָוּ
 קְדָמוּהִי—בְּמִירִי הָ—פְּרָעָמִי נִי הַמְצָרָא יִסּוּסָות—הַוּנְעַלְדָא יִתְגָּאָוּ

Between t1691a and Hebr313 at this location (green) there are legitimate differences in wording which cause some false matches. However, the negatives (gaps and mismatches) outweigh these false matches and the alignment reflects the divergence between the texts.