

## Appendix A. Automatic Evaluation

“*case+punc*” evaluation : case-sensitive, with punctuations tokenized  
“*no\_case+no\_punc*” evaluation : case-insensitive, with punctuations removed

### A.1. Official Testset (*tst2014*)

- All the sentence IDs in the IWSLT 2014 test set were used to calculate the automatic scores for each run submission.
- ASR and MT systems are ordered according to the *WER* and *BLEU* metrics, respectively.
- All automatic evaluation metric scores are given as percent figures (%).

**TED : ASR English (ASR<sub>EN</sub>)**

System	WER	(# Errors)
NICT	<b>8.4</b>	<b>(1,831)</b>
EU-BRIDGE	9.8	(2,138)
MITLL-AFRL	9.9	(2,153)
KIT	11.4	(2,475)
FBK	11.4	(2,492)
LIUM	12.3	(2,689)
UEDIN	12.7	(2,763)
IOIT	19.7	(4,283)

**TED : ASR German (ASR<sub>DE</sub>)**

System	WER	(# Errors)
KIT	<b>24.0</b>	<b>(5,660)</b>
UEDIN	35.7	(8,438)
FBK	38.8	(9,167)

**TED : ASR Italian (ASR<sub>IT</sub>)**

System	WER	(# Errors)
VECSYS-LIUM	<b>21.9</b>	<b>(5,165)</b>
MITLL-AFRL	23.0	(5,440)
FBK	23.8	(5,618)
KIT	25.4	(5,997)

**TED : SLT English-French (SLT<sub>EnFr</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
KIT	<b>27.45</b>	57.80	<b>28.16</b>	56.87
RWTH	26.94	57.29	27.74	<b>56.22</b>
LIUM	26.82	59.03	27.85	57.69
UEDIN	25.50	<b>57.23</b>	26.26	56.24
FBK	25.39	59.53	26.11	58.57
LIMSI	25.18	60.70	25.88	59.69
USFD	23.45	59.94	24.14	58.97

**TED : SLT English-German (SLT<sub>EnDe</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
KIT	<b>17.05</b>	<b>68.01</b>	<b>17.58</b>	<b>66.97</b>
UEDIN	17.00	68.36	17.51	67.30
USFD	14.75	70.15	15.24	69.15
KLE	13.00	71.70	13.64	70.33

**TED : SLT German-English (SLT<sub>DeEn</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
EU-BRIDGE	<b>19.09</b>	<b>63.80</b>	<b>19.59</b>	<b>62.94</b>
KIT	18.34	63.91	18.85	62.99
UEDIN	17.67	66.04	18.18	65.12
RWTH	17.24	65.04	17.78	64.07
KLE	9.95	74.05	10.36	72.97

**TED : MT English-French (MT<sub>EnFr</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
EU-BRIDGE	<b>36.99</b>	45.20	<b>37.85</b>	<b>44.32</b>
KIT	36.22	<b>45.18</b>	36.97	44.37
UEDIN	35.91	45.78	36.64	45.04
RWTH	35.72	44.54	36.46	43.77
MITLL-AFRL	35.48	45.69	36.90	44.49
FBK	34.24	46.75	34.85	46.04
BASELINE	30.55	49.66	31.13	49.00
MIRAACL	25.86	54.16	26.97	53.02
SFAX	16.09	62.89	17.33	61.48

**TED : MT English-German (MT<sub>EnDe</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
EU-BRIDGE	<b>23.25</b>	<b>57.27</b>	<b>24.06</b>	<b>56.15</b>
KIT	22.66	57.70	23.35	56.66
UEDIN	22.61	58.95	23.14	57.92
NTT-NAIST	22.09	57.60	22.63	56.65
KLE	19.26	61.36	19.75	60.48
BASELINE	18.44	61.89	18.92	61.02

**TED : MT English-Arabic (MT<sub>EnAr</sub>)**

System	BLEU	TER
UEDIN	<b>13.24</b>	<b>69.16</b>
KIT	13.05	71.62
BASELINE	11.12	72.88

**TED : MT English-Spanish (MT<sub>EnEs</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
UEDIN	<b>35.63</b>	<b>45.10</b>	<b>36.47</b>	<b>44.12</b>
BASELINE	31.26	48.43	31.95	47.48

**TED : MT English-Farsi (MT<sub>EnFa</sub>)**

System	BLEU	TER
BASELINE	<b>6.48</b>	<b>81.14</b>

**TED : MT English-Hebrew (MT<sub>EnHe</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>15.69</b>	<b>65.62</b>	<b>15.69</b>	<b>65.62</b>

**TED : MT English-Polish (MT<sub>EnPl</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
PJIT	<b>16.10</b>	<b>74.82</b>	<b>16.60</b>	<b>73.64</b>
BASELINE	9.75	82.60	10.16	81.44
LIA	7.79	86.89	10.12	82.31

**TED : MT English-Portuguese (MT<sub>EnPt</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
UEDIN	<b>32.41</b>	<b>45.85</b>	<b>33.12</b>	<b>44.87</b>
BASELINE	31.32	47.06	31.97	46.19

**TED : MT German-English (SLT<sub>DeEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
EU-BRIDGE	<b>25.77</b>	<b>54.61</b>	<b>26.36</b>	<b>53.76</b>
RWTH	25.04	55.49	25.61	54.65
KIT	24.62	55.62	25.16	54.77
NTT-NAIST	23.77	56.43	24.52	55.49
UEDIN	23.32	57.50	24.06	56.55
FBK	20.52	63.37	21.77	60.66
KLE	19.31	63.88	20.60	61.38
BASELINE	17.50	65.56	18.61	63.08

**TED : MT Arabic-English (MT<sub>ArEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
MITLL-AFRL	<b>27.52</b>	<b>54.54</b>	<b>28.41</b>	<b>53.44</b>
UEDIN	25.46	57.07	26.22	56.02
BASELINE	19.88	63.30	20.48	62.31

**TED : MT Spanish-English (MT<sub>EsEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
UEDIN	<b>37.29</b>	<b>43.73</b>	<b>38.07</b>	<b>42.85</b>
BASELINE	33.31	46.07	33.80	45.38

**TED : MT Farsi-English (MT<sub>FaEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
MITLL-AFRL	<b>18.37</b>	<b>66.02</b>	<b>19.03</b>	<b>65.03</b>
UEDIN	16.94	72.66	17.52	71.66
BASELINE	16.22	72.13	16.72	71.05

**TED : MT Hebrew-English (MT<sub>HeEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
UEDIN	<b>26.58</b>	<b>56.99</b>	<b>27.14</b>	<b>56.25</b>
BASELINE	23.66	58.66	24.20	57.83

**TED : MT Polish-English (MT<sub>PlEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
PJIT	<b>18.33</b>	<b>65.60</b>	<b>18.96</b>	<b>64.59</b>
BASELINE	13.94	68.75	14.49	67.63

**TED : MT Portuguese-English (MT<sub>PtEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>35.78</b>	<b>43.13</b>	<b>36.16</b>	<b>42.61</b>
UEDIN	34.66	46.11	35.28	45.52

**TED : MT English-Russian(MT<sub>EnRu</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>11.21</b>	<b>73.15</b>	<b>11.21</b>	<b>72.24</b>

**TED : MT English-Slovenian(MT<sub>EnSl</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
LIA	<b>10.36</b>	<b>71.81</b>	<b>12.69</b>	<b>67.80</b>
BASELINE	8.53	73.75	8.87	72.76

**TED : MT English-Turkish(MT<sub>EnTr</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>6.97</b>	<b>79.93</b>	<b>7.36</b>	<b>78.65</b>
UMONTREAL	4.76	80.67	5.51	79.28

**TED : MT English-Chinese(MT<sub>EnZh</sub>)**

System	character-based	
	BLEU	TER
USTC	21.64	65.71
KIT	18.31	66.43
HKUST	16.41	74.35
BASELINE	15.56	80.48
UMONTREAL	7.40	81.89

**TED : MT Russian-English (MT<sub>RuEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
MITLL-AFRL	<b>19.30</b>	<b>63.95</b>	<b>20.22</b>	<b>62.64</b>
BASELINE	15.48	69.93	15.95	68.91

**TED : MT Slovenian-English (MT<sub>SlEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>13.69</b>	<b>70.79</b>	<b>14.07</b>	<b>69.83</b>

**TED : MT Turkish-English (MT<sub>TrEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>12.52</b>	<b>76.96</b>	<b>13.10</b>	<b>75.77</b>

**TED : MT Chinese-English (MT<sub>ZhEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
USTC	<b>15.65</b>	<b>69.65</b>	<b>16.35</b>	<b>68.62</b>
NICT	14.05	71.68	14.88	70.42
MITLL-AFRL	12.83	74.74	13.51	73.58
BASELINE	11.22	72.43	11.79	71.37
HKUST	9.64	76.67	10.83	74.16

## A.2. Progress Test Set (*tst2013*)

- All the sentence IDs in the IWSLT 2013 test set were used to calculate the automatic scores for each run submission.
- ASR and MT systems are ordered according to the *WER* and *BLEU* metrics, respectively.
- For each task, the best score of each metric is marked with **boldface**.
- All automatic evaluation metric scores are given as percent figures (%).

**TED: ASR English *tst2013***

System	IWSLT 2013		IWSLT 2014	
	WER	(# Errors)	WER	(# Errors)
NICT	<b>13.5</b>	<b>(5,734)</b>	<b>10.6</b>	<b>(4,518)</b>
MITLL-AFRL	15.9	(6,788)	13.7	(5,856)
KIT	14.4	(6,115)	14.2	(6,044)
FBK	23.2	(9,899)	14.7	(6,247)
LIUM	—	—	16.0	(6,818)
UEDIN	22.1	(9,413)	16.3	(6,963)
IOIT	27.2	(11,578)	24.0	(10,206)

**TED: ASR German *tst2013***

System	IWSLT 2013		IWSLT 2014	
	WER	(# Errors)	WER	(# Errors)
KIT	<b>25.7</b>	<b>(4,932)</b>	<b>25.4</b>	<b>(5,885)</b>
UEDIN	37.8	(7,250)	35.0	(6,720)
FBK	37.5	(7,199)	37.8	(7,261)

**TED : MT English-French test 2013(MT<sub>EnFr</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
EU-BRIDGE	<b>40.50</b>	43.27	<b>41.65</b>	42.06
KIT	40.12	43.09	41.11	42.04
RWTH	39.72	<b>42.58</b>	40.73	<b>41.52</b>
UEDIN	39.59	43.80	40.45	42.78
MITLL-AFRL	39.08	44.05	40.59	42.73
FBK	38.20	44.83	38.99	43.88
BASELINE	33.20	48.91	33.81	48.07
MIRAFL	29.63	51.96	30.91	50.65

**TED : MT English-German test 2013 (MT<sub>EnDe</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
EU-BRIDGE	<b>26.22</b>	<b>53.62</b>	<b>27.30</b>	<b>52.34</b>
KIT	26.03	53.83	26.77	52.81
NTT-NAIST	25.80	53.86	26.55	52.75
UEDIN	25.33	55.12	26.13	53.93
KLE	21.69	58.27	22.25	57.32
BASELINE	20.96	58.48	21.52	57.58

**TED : MT German-English test 2013 (MT<sub>DeEn</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
EU-BRIDGE	<b>28.77</b>	<b>50.52</b>	<b>29.29</b>	<b>49.63</b>
KIT	27.98	50.92	28.55	50.04
NTT-NAIST	27.81	51.62	28.32	50.82
UEDIN	27.60	52.43	28.26	51.44
RWTH	27.59	51.33	28.08	50.41
FBK	25.45	55.80	26.07	54.88
KLE	23.59	57.38	24.18	56.47
BASELINE	20.26	60.33	20.89	59.48

**TED : MT English-Arabic test 2013(MT<sub>EnAr</sub>)**

System	BLEU	TER
UEDIN	<b>14.20</b>	<b>65.97</b>
KIT	14.15	68.29
BASELINE	12.68	68.94

**TED : MT Arabic-English test 2013 (MT<sub>ArEn</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
MITLL-AFRL	<b>31.48</b>	<b>49.88</b>	<b>32.41</b>	<b>48.76</b>
UEDIN	29.06	53.02	29.74	52.03
BASELINE	21.63	60.32	22.46	59.12

**TED : MT English-Spanish test 2013 (MT<sub>EnEs</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
UEDIN	<b>34.74</b>	<b>45.75</b>	<b>35.42</b>	<b>44.78</b>
BASELINE	30.63	49.39	31.14	48.57

**TED : MT Spanish-English test 2013(MT<sub>EsEn</sub>)**

System	<i>case sensitive</i>		<i>case insensitive</i>	
	BLEU	TER	BLEU	TER
UEDIN	<b>39.13</b>	<b>41.37</b>	<b>39.75</b>	<b>40.60</b>
BASELINE	34.18	44.63	34.70	44.00

**TED : MT English-Farsi test 2013 (MT<sub>EnFa</sub>)**

System	BLEU	TER
BASELINE	7.05	78.90

**TED : MT English-Hebrew test 2013(MT<sub>EnHe</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>15.92</b>	<b>64.16</b>	<b>15.92</b>	<b>64.16</b>

**TED : MT English-Polish test2013 (MT<sub>EnPl</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
PJIT	<b>25.92</b>	<b>61.04</b>	<b>26.62</b>	<b>59.94</b>
BASELINE	11.12	75.95	11.67	74.78

**TED : MT English-Portuguese test 2013(MT<sub>EnPt</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	31.38	46.42	31.89	45.66
UEDIN	33.20	44.90	33.93	43.90

**TED : MT English-Russian test 2013(MT<sub>EnRu</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>14.01</b>	<b>70.47</b>	<b>14.01</b>	<b>69.44</b>

**TED : MT English-Slovenian test 2013 (MT<sub>EnSl</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>9.63</b>	<b>73.32</b>	<b>9.97</b>	<b>72.34</b>

**TED : MT English-Turkish test 2013 (MT<sub>EnTr</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>6.85</b>	<b>80.40</b>	<b>7.21</b>	<b>79.08</b>
UMONTREAL	4.06	83.97	4.77	82.50

**TED : MT English-Chinese test2013 (MT<sub>EnZh</sub>)**

System	character-based	
	BLEU	TER
USTC	<b>22.49</b>	<b>63.74</b>
KIT	21.01	63.12
HKUST	18.81	70.94
BASELINE	18.23	76.15
UMONTREAL	7.93	80.47

**TED : MT Farsi-English test 2013 (MT<sub>FaEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
MITLL-AFRL	<b>19.47</b>	<b>63.27</b>	<b>20.11</b>	<b>62.27</b>
UEDIN	16.51	82.50	16.87	81.58
BASELINE	14.04	83.01	14.44	82.09

**TED : MT Hebrew-English test2013 (MT<sub>HeEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
UEDIN	<b>29.70</b>	<b>52.40</b>	<b>30.51</b>	<b>51.35</b>
BASELINE	25.97	55.40	26.74	54.23

**TED : MT Polish-English test2013 (MT<sub>PlEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
PJIT	<b>27.99</b>	<b>58.01</b>	<b>28.61</b>	<b>57.10</b>
BASELINE	17.25	66.44	17.75	65.44

**TED : MT Portuguese-English test 2013 (MT<sub>PtEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	37.85	40.87	38.26	40.35
UEDIN	37.34	42.91	37.80	42.30

**TED : MT Russian-English test 2012 (MT<sub>RuEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
MITLL-AFRL	<b>24.30</b>	<b>57.59</b>	<b>25.39</b>	<b>56.25</b>
BASELINE	19.82	63.56	20.40	62.46

**TED : MT Slovenian-English test2013 (MT<sub>SlEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>14.64</b>	<b>68.68</b>	<b>15.19</b>	<b>67.63</b>

**TED : MT Turkish-English test 2013 (MT<sub>TrEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
BASELINE	<b>13.30</b>	<b>75.17</b>	<b>13.95</b>	<b>74.00</b>

**TED : MT Chinese-English test 2013(MT<sub>ZhEn</sub>)**

System	case sensitive		case insensitive	
	BLEU	TER	BLEU	TER
USTC	<b>18.12</b>	<b>66.28</b>	<b>18.85</b>	<b>65.23</b>
NICT	16.57	67.96	17.36	66.76
MITLL-AFRL	15.59	70.89	16.32	69.68
BASELINE	13.40	68.85	14.00	67.90
HKUST	11.89	72.33	13.08	70.10