

A Savoyard Version of the FLORES+ Corpus

Malik Marmonier[†] Alain Favre[‡] Benoît Sagot[†] Rachel Bawden[†]

[†]Inria, Paris, France [‡]Institut de la langue savoyarde

{firstname.lastname}@inria.fr[†] {firstname.lastname}@langue-savoyarde.com[‡]

Abstract

We present a Savoyard version of the FLORES+ corpus, a contribution to the Open Language Data Initiative (OLDI) shared task at WMT 2026. Savoyard, an endangered Arpitan dialect of the Gallo-Romance language family, is currently undergoing active revitalization. We give an overview of its linguistic history, describe our data collection process, and report baseline scores for a range of systems on the resulting French↔Savoyard evaluation set as a point of comparison for future work. We hope that this dataset will stimulate further data collection efforts for Savoyard and the broader Arpitan dialect continuum, supporting the development of NLP resources for these endangered languages and contributing to the preservation of humanity’s linguistic heritage.

1 Introduction

Machine Translation (MT) has advanced considerably over the past four decades (Brown et al., 1990; Koehn et al., 2003; Sutskever et al., 2014; Bahdanau et al., 2015; Vaswani et al., 2017). This progress owes much to the field’s departure from evaluation practices based on nebulous, self-reported notions of linguistic coverage in favor of shared, empirical evaluation frameworks (Liberman, 2010). Yet these advances have primarily benefited a handful of dominant languages, leaving lower-resourced varieties largely excluded (Joshi et al., 2020; Haddow et al., 2022). Addressing this disparity requires extending the same empirical evaluation rigor that drove earlier advances to low-resource settings.

It is explicitly to this end that the FLORES+ dataset was initially compiled (Goyal et al., 2022, as Flores-101) and continuously expanded (NLLB Team, 2024; Gordeev et al., 2024; Kuzhuget et al., 2024; Ali et al., 2024; Abdulmumin et al., 2024; Perez-Ortiz et al., 2024; Mæhlum et al., 2025; Okttem et al., 2025; Frontull et al., 2025; Lou et al.,

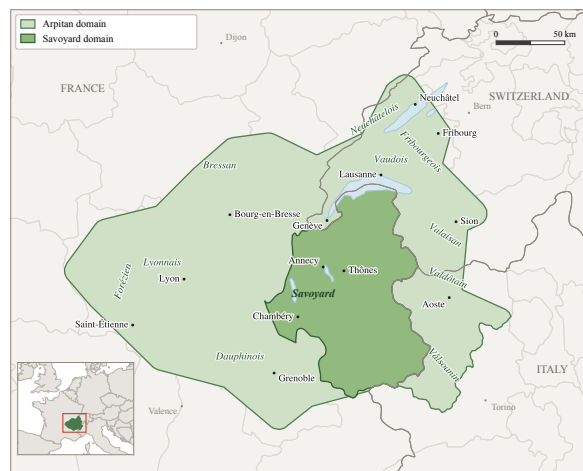


Figure 1: The Arpitan dialect continuum, with the Savoyard domain in a darker green. Outer boundaries are schematic renderings based on (Kristol, 2016) and French Wikipedia maps; political borders and hydrography are drawn from Natural Earth.

2025; Samil and Adelani, 2026, *inter alia*), establishing a standardized evaluation framework for more than 200 languages and script variants, with a deliberate focus on otherwise underrepresented languages.

As a direct extension of these past efforts, our work presents a version of the FLORES+ corpus in the Savoyard language, an endangered Arpitan dialect of the Gallo-Romance family currently the object of active revitalization efforts. In the following sections, we provide an overview of the linguistic history of Savoyard, present our data collection process, and report various baseline performance scores on the resulting evaluation set.

We hope that this dataset will stimulate further data collection efforts for Savoyard and the broader Arpitan dialect continuum, ultimately supporting the development of NLP resources to aid in the preservation of these endangered languages.



Figure 2: Lake Annecy, in the historical heart of the Savoyard domain. Photo by Alex Brown, CC BY 2.0.

2 Language Overview

2.1 What’s in a name: Arpitan, Franco-Provençal, or Francoprovençal?

Latin, implanted across Gaul in the wake of the Roman conquest, gave rise to a family of Romance vernaculars spanning modern France and Belgium, together with parts of Switzerland and northern Italy. These “Gallo-Romance” languages were long divided into two branches along a north–south axis, named for their medieval reflexes of Latin *hoc* (‘yes’): the *langues d’oïl* of the north (Smith, 2016), chief among them French, and the *langues d’oc*, or Occitan, of the south (Olivieri and Sauzet, 2016)—echoing a classification of the Romance vernaculars famously articulated by Dante in the early 14th century (Alighieri, 1996). A third grouping went unrecognized until Ascoli (1878) demonstrated, on phonetic grounds, that the dialects of east-central France, French-speaking Switzerland, and the Aosta Valley of Italy form a coherent group, neither varieties of an Oc or Oïl language nor a mixture of both, but constituting a distinct linguistic development characterized by its own treatment of the Latin substratum.

Ascoli baptized this new grouping *franco-provenzale* (‘Franco-Provençal’), a hyphenated name that—in yoking together the labels of an Oïl and an Oc language—seemed to some to announce precisely the hybrid status his analysis had refuted, and the label has therefore been faulted ever since (Tuaiillon, 1983). In an attempt to mitigate this perceived misnomer, later scholars suppressed the hyphen, *Francoprovençal*, recasting the two elements as a unified linguistic sign (Tuaiillon, 1983)—a spelling now standard in the scientific literature (Kristol, 2016). Nonetheless, this solution was deemed unsatisfactory by language activists who promoted the neo-endonym “Arpitan”

as a replacement (from a local root *arp-* ‘mountain pasture’).¹ Once contentious for its separatist and ethno-nationalist overtones in the Italian context that birthed it (Bichurina, 2019), the term “Arpitan” has since been largely depoliticized and is seeing growing institutional adoption. It has now supplanted *Franco-Provençal* and *Francoprovençal* as the reference name of the language under SIL.² It is also the term preferred by the *Institut de la langue savoyarde*—whose translation services produced the FLORES+ partition this paper presents—to refer to the broader language grouping to which Savoyard belongs.

2.2 Savoyard

Savoyard³ is a subset of the Arpitan dialect continuum. It denotes the Arpitan varieties historically spoken in the County of Savoy—a feudal state of the Holy Roman Empire founded in 1003 and raised to a duchy in 1416, which straddled the modern French–Italian–Swiss alpine borders, with its Arpitan-speaking core comprising the present-day French *départements* of *Savoie* and *Haute-Savoie*, and most of the Swiss Canton of Geneva.⁴

As Latin receded from administrative and literary use in the late European Middle Ages, Romance vernaculars—such as French or the various dialects of the Italian city-states—moved into the registers it vacated. Arpitan did so outside Savoy, to a limited extent.⁵ Savoyard, however, was deprived

¹As is often the case with stigmatized vernaculars in French-speaking diglossic contexts, the actual endonym is the pejorative term “*patois/patouè*”.

²<https://iso639-3.sil.org/code/frp>; we note that ISO 639-3 code is still frp.

³Glottocode *savo1253*. We follow Glottolog in using “Savoyard,” though we note that the Institut de la langue savoyarde prefers “Savoyan” as the English name of the language.

⁴Excluding the city of Geneva itself.

⁵For instance, in the Lyonnais, in the devotional writings of

of such an opportunity due to the Duchy's early adoption of French as the language of its court and chancery.⁶ It was therefore left without the administrative or literary functions through which other vernaculars acquired prestige.

The prestige of Savoyard nonetheless improved markedly in the 19th century, when an abundant literature written in Savoyard began to emerge, prompted by Romantic interest in folklore and popular tradition (Bichurina and Dunoyer, 2021), with poet Amélie Gex (1835–1883) as its most accomplished representative. This development, however, was cut short by the annexation of Savoy to France in 1860. Aggressive French nation-building policies had long identified linguistic diversity with anti-republican feudalism (Grégoire, 1794), and the vernaculars—now uniformly denigrated as *patois*—were driven out of public life across France. In Savoy, this process was particularly virulent, compressed as it was into a shorter time span due to the region's more recent inclusion into the French polity, and the school system was its principal instrument: children were forbidden to speak Savoyard, humiliated and punished when they did, translating political assimilation into practices of *de facto* linguistic violence (Skutnabb-Kangas, 2000).

Revitalization efforts began in earnest in the second half of the 20th century, as the vocabulary of decolonization was taken up by minority-language communities in the First World itself, and gathered pace around the turn of the millennium, when the framing of language loss in terms borrowed from environmental conservation entered French public debate and inflected public policy (Bichurina, 2019).

A late-1980s estimate placed the number of Savoyard speakers at approximately 30,000 (Tuailon, 1989), though reliable counts remain difficult to establish (Boutillier, 2022). Indeed, recent sociolinguistic work—the FORA survey (Bert and Costa, 2009) and subsequent fieldwork by Bichurina and Dunoyer (2021)—reveals a complex linguistic landscape characterized by four main categories of speakers. The “native” or traditional speakers, generally aged between 80 and 90 years old, acquired the language in early childhood but

their use of Savoyard often remains invisible in the public sphere due to past stigmatization. In contrast, “late” speakers—predominantly educated retirees in their 60s and 70s who had limited exposure to Savoyard in their childhood but re-engaged with the language in adulthood—are highly visible and drive preservation efforts through dedicated associations. Alongside these, a small but growing group of “new” speakers, primarily teenagers and young adults, approach Savoyard as a learned language, driven by broader cultural motives. Finally, the landscape includes a vast pool of “passive speakers” who comprehend the language but struggle to express themselves in it. Because these individuals can rapidly transition into active speakers within supportive environments, their fluctuating proficiency complicates attempts to establish static speaker counts.

While a comprehensive linguistic description of Savoyard falls outside the scope of this paper,⁷ its defining characteristic—marked dialectal fragmentation—is highly relevant to corpus building. As an Arpitan variety, Savoyard exhibits significant micro-linguistic diversity, exacerbated over centuries by the geographic isolation of alpine valleys. This diversity extends beyond lexical variations to encompass morphological and phonetic divergences, where the evolution of shared Latin roots often differs from one village to the next. In a computational context, this significant local variance presents a unique modeling challenge, underscoring the necessity of a supra-dialectal orthography capable of unifying these diverse variants.

2.3 Revitalization and Written Standards

Savoyard revitalization efforts rest mainly on networks of private non-profit associations: the local *groupes patoisants* and, since 2002, the *Institut de la langue savoyarde*.⁸ A persistent barrier to the revitalization of Savoyard—and Arpitan more broadly—through French public education stemmed from the absence of a unified orthography. Despite a softening stance toward regional languages, the French state's ideological framework still equates linguistic legitimacy with a codified written standard (Bichurina and Dunoyer, 2021). This text-centric ideology historically underpinned the *loi Deixonne* (1951)—the foundational law for the teaching of regional languages in

the mystic Marguerite d'Oingt (d. 1310), and in the Dauphiné, where Justinian's *Institutes* were rendered into the dialect of Grenoble (Bichurina and Dunoyer, 2021).

⁶An early adoption that predated even that of the French Kingdom. The neighboring Republic of Geneva replaced Latin with French on a similar timeline, and to the equal detriment of the local Arpitan vernacular.

⁷See (Désormaux, 1981) and (Viret, 2015) for detailed accounts and additional references.

⁸In Savoyard: *Enstitut de la Lengoua Savoyârda*.

France—which systematically excluded uncoded vernaculars.

Defining a written standard for Arpitan thus became a strategic imperative for this linguistic community, albeit a deeply divisive one. For decades, the dominant written form in Savoy was the *Graphie de Conflans* (Tuailon, 1984). Developed in the 1980s by linguists and native speakers, this highly phonetic system was designed to faithfully transcribe the exact pronunciation of local village dialects. While it preserved the region’s immense micro-linguistic diversity, its hyper-local fidelity severely hindered broader intercomprehension and official recognition. To bridge this gap, linguist Dominique Stich introduced the *Orthographe de Référence B* (ORB) in the early 2000s (Stich, 2001). As a supra-dialectal, etymological system, ORB aims to abstract local phonetic and morphological variations into a single graphemic standard for the entire Arpitan domain.

While ORB initially faced fierce resistance from traditional speakers—who found it harder to master and feared it erased their specific local heritage—it was championed by “new” speakers and revitalization activists for its unifying character and political leverage (Bichurina, 2019; Dunoyer, 2019). Ultimately, the strategic necessity of a unified standard appears to have prevailed, and it is in great part thanks to the growing acceptance of the principle of a supra-dialectal written standard that the French Ministry of Education officially authorized the teaching of Arpitan in French public schools in 2021 (Ministère de l’Éducation nationale, de la Jeunesse et des Sports, 2021), with the active collaboration of the *Institut de la langue savoyarde*.

3 Data Collection

This section outlines the data collection process for the Savoyard partitions of the FLORES+ dev and devtest sets, which were translated from French by the translation service of the *Institut de la langue savoyarde* (ILS).

3.1 Target Linguistic Variant and Orthography

The Savoyard variant selected as the target language for the partition is termed *Savoyard médian* (Median Savoyard) by the ILS. It is the emergent standard that the ILS teaches through MOOCs and online seminars. It is closest to the dialect of Thônes, a town located near the geographic center

of Savoy, and therefore broadly understandable by the greatest number of speakers, though its emergence as a written standard is mostly tied to lexicographical efforts by Favre and Stich (2005), recently expanded by Favre (2026). The orthographic norm selected for the partition is the aforementioned supra-dialectal ORB spelling as documented in (Stich, 2003).

3.2 Translation Process

The translation workload, initially shared between two translators, was ultimately borne almost entirely by Alain Favre, a proficient late speaker and lexicographer who contributed in the past to the development of a rule-based Apertium system for the French-Arpitan language pair (Forcada and Tyers, 2016; Alòs, 2020).

Translators were instructed to follow the official OLDI guidelines,⁹ which were presented to them during an online meeting. They worked within a customized spreadsheet interface (see Figure 3), with one sheet for each of the dev and devtest partitions. A source text column displayed the French segments of FLORES+, originally extracted from English Wikinews, Wikijunior, and Wikivoyage documents and translated into French by the original compilers of the corpus (Goyal et al., 2022). A column displaying the URL of the source documents allowed the translators to refer to them, but since these documents are in English, a column displaying the Italian translation of the segments proved more useful for resolving ambiguities in the French text. While the translators were instructed to translate the devtest partition strictly from scratch, for the dev partition they were provided with translation hypotheses generated by the rule-based French-Arpitan Apertium system (Forcada and Tyers, 2016) for inspiration (column not shown in Figure 3). This design choice was motivated by the desire to collect in-context-learning or finetuning data for the automatic post-editing of Apertium output into Savoyard.¹⁰ Such data would in turn facilitate large-scale data augmentation.

The translators were given three and a half months to complete the assigned task and were compensated at a fair rate. During their work, they faced diverse challenges, both stylistic (the formal “press release” style typical of Wikinews source texts, for instance, does not historically exist in

⁹<https://oldi.org/translation-guidelines.pdf>

¹⁰This design choice was discussed with and approved by the OLDI shared task organizers beforehand.

FLORES+ - Arpitan Savoyard									
Fichier Édition Affichage Insertion Format Données Outils Gemini Extensions Aide FLORES Export									
Q Menus 100% 123 Par défaut 10 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619 1620 1621 1622 1623 1624 1625 1626 1627 1628 1629 1630 1631 1632 1633 1634 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649 1650 1651 1652 1653 1654 1655 1656 1657 1658 1659 1660 1661 1662 1663 1664 1665 1666 1667 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1680 1681 1682 1683 1684 1685 1686 1687 1688 1689 1690 1691 1692 1693 1694 1695 1696 1697 1698 1699 1700 1701 1702 1703 1704 1705 1706 1707 1708 1709 1710 1711 1712 1713 1714 1715 1716 1717 1718 1719 1720 1721 1722 1723 1724 1725 1726 1727 1728 1729 1730 1731 1732 1733 1734 1735 1736 1737 1738 1739 1740 1741 1742 1743 1744 1745 1746 1747 1748 1749 1750 1751 1752 1753 1754 1755 1756 1757 1758 1759 1760 1761 1762 1763 1764 1765 1766 1767 1768 1769 1770 1771 1772 1773 1774 1775 1776 1777 1778 1779 1780 1781 1782 1783 1784 1785 1786 1787 1788 1789 1790 1791 1792 1793 1794 1795 1796 1797 1798 1799 1800 1801 1802 1803 1804 1805 1806 1807 1808 1809 1810 1811 1812 1813 1814 1815 1816 1817 1818 1819 1820 1821 1822 1823 1824 1825 1826 1827 1828 1829 1830 1831 1832 1833 1834 1835 1836 1837 1838 1839 1840 1841 1842 1843 1844 1845 1846 1847 1848 1849 1850 1851 1852 1853 1854 1855 1856 1857 1858 1859 1860 1861 1862 1863 1864 1865 1866 1867 1868 1869 1870 1871 1872 1873 1874 1875 1876 1877 1878 1879 1880 1881 1882 1883 1884 1885 1886 1887 1888 1889 1890 1891 1892 1893 1894 1895 1896 1897 1898 1899 1900 1901 1902 1903 1904 1905 1906 1907 1908 1909 1910 1911 1912 1913 1914 1915 1916 1917 1918 1919 1920 1921 1922 1923 1924 1925 1926 1927 1928 1929 1930 1931 1932 1933 1934 1935 1936 1937 1938 1939 1940 1941 1942 1943 1944 1945 1946 1947 1948 1949 1950 1951 1952 1953 1954 1955 1956 1957 1958 1959 1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2031 2032 2033 2034 2035 2036 2037 2038 2039 2040 2041 2042 2043 2044 2045 2046 2047 2048 2049 2050 2051 2052 2053 2054 2055 2056 2057 2058 2059 2060 2061 2062 2063 2064 2065 2066 2067 2068 2069 2070 2071 2072 2073 2074 2075 2076 2077 2078 2079 2080 2081 2082 2083 2084 2085 2086 2087 2088 2089 2090 2091 2092 2093 2094 2095 2096 2097 2098 2099 2100 2101 2102 2103 2104 2105 2106 2107 2108 2109 2110 2111 2112 2113 2114 2115 2116 2117 2118 2119 2120 2121 2122 2123 2124 2125 2126 2127 2128 2129 2130 2131 2132 2133 2134 2135 2136 2137 2138 2139 2140 2141 2142 2143 2144 2145 2146 2147 2148 2149 2150 2151 2152 2153 2154 2155 2156 2157 2158 2159 2160 2161 2162 2163 2164 2165 2166 2167 2168 2169 2170 2171 2172 2173 2174 2175 2176 2177 2178 2179 2180 2181 2182 2183 2184 2185 2186 2187 2188 2189 2190 2191 2192 2193 2194 2195 2196 2197 2198 2199 2200 2201 2202 2203 2204 2205 2206 2207 2208 2209 2210 2211 2212 2213 2214 2215 2216 2217 2218 2219 2220 2221 2222 2223 2224 2225 2226 2227 2228 2229 2230 2231 2232 2233 2234 2235 2236 2237 2238 2239 2240 2241 2242 2243 2244 2245 2246 2247 2248 2249 2250 2251 2252 2253 2254 2255 2256 2257 2258 2259 2260 2261 2262 2263 2264 2265 2266 2267 2268 2269 2270 2271 2272 2273 2274 2275 2276 2277 2278 2279 2280 2281 2282 2283 2284 2285 2286 2287 2288 2289 2290 2291 2292 2293 2294 2295 2296 2297 2298 2299 2300 2301 2302 2303 2304 2305 2306 2307 2308 2309 2310 2311 2312 2313 2314 2315 2316 2317 2318 2319 2320 2321 2322 2323 2324 2325 2326 2327 2328 2329 2330 2331 2332 2333 2334 2335 2336 2337 2338 2339 2340 2341 2342 2343 2344 2345 2346 2347 2348 2349 2350 2351 2352 2353 2354 2355 2356 2357 2358 2359 2360 2361 2362 2363 2364 2365 2366 2367 2368 2369 2370 2371 2372 2373 2374 2375 2376 2377 2378 2379 2380 2381 2382 2383 2384 2385 2386 2387 2388 2389 2390 2391 2392 2393 2394 2395 2396 2397 2398 2399 2400 2401 2402 2403 2404 2405 2406 2407 2408 2409 2410 2411 2412 2413 2414 2415 2416 2417 2418 2419 2420 2421 2422 2423 2424 2425 2426 2427 2428 2429 2430 2431 2432 2433 2434 2435 2436 2437 2438 2439 2440 2441 2442 2443 2444 2445 2446 2447 2448 2449 2450 2451 2452 2453 2454 2455 2456 2457 2458 2459 2460 2461 2462 2463 2464 2465 2466 2467 2468 2469 2470 2471 2472 2473 2474 2475 2476 2477 2478 2479 2480 2481 2482 2483 2484 2485 2486 2487 2488 2489 2490 2491 2492 2493 2494 2495 2496 2497 2498 2499 2500 2501 2502 2503 2504 2505 2506 2507 2508 2509 2510 2511 2512 2513 2514 2515 2516 2517 2518 2519 2520 2521 2522 2523 2524 2525 2526 2527 2528 2529 2530 2531 2532 2533 2534 2535 2536 2537 2538 2539 2540 2541 2542 2543 2544 2545 2546 2547 2548 2549 2550 2551 2552 2553 2554 2555 2556 2557 2558 2559 2560 2561 2562 2563 2564 2565 2566 2567 2568 2569 2570 2571 2572 2573 2574 2575 2576 2577 2578 2579 2580 2581 2582 2583 2584 2585 2586 2587 2588 2589 2590 2591 2592 2593 2594 2595 2596 2597 2598 2599 2600 2601 2602 2603 2604 2605 2606 2607 2608 2609 2610 2611 2612 2613 2614 2615 2616 2617 2618 2619 2620 2621 2622 2623 2624 2625 2626 2627 2628 2629 2630 2631 2632 2633 2634 2635 2636 2637 2638 2639 2640 2641 2642 2643 2644 2645 2646 2647 2648 2649 2650 2651 2652 2653 2654 2655 2656 2657 2658 2659 2660 2661 2662 2663 2664 2665 2666 2667 2668 2669 2670 2671									

System	Modality	FR → FRP				FRP → FR			
		Extr. %	BLEU ↑	chrF++ ↑	TER ↓	Extr. %	BLEU ↑	chrF++ ↑	TER ↓
Copy Source Baseline	-	100.00	10.5	40.2	70.1	100.00	10.5	39.3	70.2
Apertium (Rule-Based)	Standard Generation	100.00	74.7	86.0	15.3	100.00	68.8	81.4	19.8
NLLB-200 (600M)	Finetuned Baseline (Gold)	100.00	35.0	62.9	39.5	100.00	63.2	78.3	23.8
	Distilled (Gold+Silver)	100.00	69.6	83.3	18.1	100.00	72.4	84.2	17.1
Gemma-4-31B	Zero-Shot Direct (No-Think)	100.00	12.3	42.6	65.4	100.00	71.4	82.9	18.6
	Zero-Shot Reasoning	90.32	11.7	39.6	67.5	98.72	71.9	82.8	19.0
	Few-Shot Direct (No-Think)	100.00	58.4	77.5	24.5	100.00	77.3	86.6	14.4
	Few-Shot Reasoning	82.61	45.7	64.8	39.2	89.92	69.6	79.7	22.7
Gemma-4-E4B	Zero-Shot Direct (No-Think)	100.00	9.8	39.4	70.8	100.00	61.2	76.1	26.3
	Zero-Shot Reasoning	100.00	8.8	37.4	72.8	100.00	58.3	74.5	29.9
	Few-Shot Direct (No-Think)	100.00	40.2	66.0	38.1	100.00	65.5	79.0	23.2
	Few-Shot Reasoning	100.00	33.4	60.8	43.6	100.00	62.5	77.1	26.1
Gemma-4-E2B	Zero-Shot Direct (No-Think)	100.00	8.9	38.5	72.7	100.00	54.9	72.2	30.8
	Zero-Shot Reasoning	100.00	9.0	37.2	74.9	100.00	33.7	67.0	63.1
	Few-Shot Direct (No-Think)	100.00	27.1	55.4	50.3	100.00	53.0	69.3	33.9
	Few-Shot Reasoning	100.00	4.2	38.5	161.5	100.00	29.2	64.5	68.4

Table 1: Evaluation metrics for both translation directions (FR → FRP and FRP → FR).

ment (French in the FR→FRP, Savoyard in the FRP→FR) unmodified as the translation hypothesis under evaluation. Despite the close phylogenetic relationship and ongoing diglossic contact between French and Savoyard, this baseline performs poorly in both directions (BLEU 10.5, chrF++ 40.2/39.3), indicating that the two languages display only marginal overlap in their surface forms and that the FLORES+ test partition meaningfully discriminates between systems with genuine translation ability and mere copying.

Rule-based system We next evaluate the French–Arpitan Apertium rule-based system (Forcada and Tyers, 2016; Alòs, 2020),¹⁴ which performs strongly in both directions, and in fact constitutes the best-performing system for FR→FRP (BLEU 74.7, chrF++ 86.0, TER 15.3). While the Arpitan Apertium system was not used by the translators to translate the devtest partition of FLORES+, one of the translators was instrumental in its development, as previously mentioned, and its output is therefore likely to closely mirror this translator’s idiolect. Apertium performance is, however, asymmetric: it is comparatively weaker for FRP→FR (BLEU 68.8, chrF++ 81.4, TER 19.8), a pattern we return to below.

NMT finetuning We finetune NLLB-200-600M (NLLB Team, 2024) directly on the dev partition of our Savoyard FLORES+ corpus (with 20% held out for validation), using a learning rate of 2e-5, an effective batch size of 16, weight decay of 0.01, mixed-precision training, and a best-epoch-saving strategy over 10 epochs. Given the limited size of this gold-reference training signal,

performance is considerably lower than that of Apertium, and strikingly asymmetric across directions: FRP→FR reaches BLEU 63.2 (chrF++ 78.3), whereas FR→FRP falls to BLEU 35.0 (chrF++ 62.9). This gap suggests that, under a small-data regime, generation into the lower-resourced Savoyard is substantially harder than generation into French, for which NLLB’s pretrained capacity can offset the scarcity of in-domain supervision.

NMT distillation-finetuning Our second NMT configuration, Distilled (Gold+Silver), augments the gold development data with silver parallel data generated through a two-step translation pipeline. First, we translate the French source text from the OLDI Seed (Maillard et al., 2023; NLLB Team, 2024; Marmonier et al., 2025) and SMOLSENT (Caswell et al., 2025)¹⁵ corpora into Arpitan using Apertium. Second, we use Gemma-4-31B to post-edit this Arpitan draft into Savoyard. To guide the LLM, we prompt it with few-shot triplet demonstrations consisting of the French source segments from the FLORES+ dev set, the Apertium Arpitan draft translations presented to the human translators during the annotation process, and the gold Savoyard reference from the compiled Savoyard dev partition (see Appendix A). We then finetune NLLB-200 on the concatenated gold and silver datasets. This approach substantially closes the gap with Apertium and reverses the direction-wise asymmetry observed for rule-based translation, consistent with the pattern already seen for the NLLB Baseline model above: for FRP→FR, the distilled model (BLEU 72.4, chrF++ 84.2) outperforms Apertium, while for FR→FRP it remains

¹⁴<https://github.com/apertium/apertium-fra-frp>

¹⁵We compiled a French version of SMOLSENT, which we contribute to the same 2026 OLDI Shared Task.

slightly behind (BLEU 69.6 vs. 74.7).

LLMs We also evaluate three models from the Gemma-4 family (31B, E4B, and E2B; Google DeepMind 2026) across two binary axes. The first axis, Zero-Shot vs. Few-Shot, varies the level of in-context supervision. Zero-shot models receive only a brief prompt instructing them to translate the segment based on prior knowledge, while few-shot models are supplied with the entire *dev* partition structured as ordered bitexts within their prompt. The second axis, Direct (No-Think) vs. Reasoning, controls the generation modality. Direct models must emit the translation immediately within XML tags, whereas Reasoning models are permitted an explicit internal monologue before producing the tagged output.¹⁶ See Appendix B for prompt examples.

Two patterns in LLM performance emerge from Table 1.

First, few-shot demonstrations provide substantial and consistent gains across all model sizes and translation directions. For instance, under Direct generation, Gemma-4-31B improves from 12.3 to 58.4 BLEU for FR→FRP, and from 71.4 to 77.3 BLEU for FRP→FR; identical trends hold for E4B and E2B. This indicates that the *dev* partition of FLORES+, despite its modest size, offers a potent in-context-learning signal for a genuinely low-resource variety that is likely absent from pretraining data while closely related to a high-resource language.

Second, while standard model ranking remains constant across modalities (Gemma-4-31B > E4B > E2B), allowing models to reason before translating is universally harmful, with the performance penalty exhibiting a non-monotonic relationship to model size. Performance degrades most sharply at both ends of the model-size spectrum, though through distinct failure mechanisms: the 31B variant falls into infinite self-checking loops that exhaust the token limit (dropping extraction rates to 82.61% for few-shot FR→FRP), whereas the smaller E2B model frequently emits its chain-of-thought directly within the output tags (driving TER scores above 160). The intermediate

¹⁶We use a regular expression to extract the string within the <translation>...</translation> tags. Extraction failures (reported as *Extr. %* in Table 1) are penalized as empty strings. Inference is powered by vLLM (Kwon et al., 2023) and Hugging Face transformers (Wolf et al., 2020). Max output tokens are capped at 1,024 for Direct and 16,384 for Reasoning; vLLM’s max-model-len is set to 240,000 for the few-shot settings.

E4B model incurs a milder penalty, only suffering from reasoning-induced linguistic drift, occasionally yielding Italian-inflected or French-lexified pseudo-translations (e.g., “*L’armonia al luo de travai...*”) adequately penalized by the metrics. Ultimately, no model, at any scale, appears to successfully leverage an explicit internal monologue productively for this translation task.¹⁷

5 Human Validation

To validate the automatic metrics and incidentally the quality of the human translations, we conducted a blind pairwise human evaluation campaign. The annotation was performed by Alain Favre, who also served as the primary translator for the evaluation set. To preserve statistical power while remaining mindful of the annotator’s time, we restricted the experiment to a subset of the most promising systems: the rule-based Apertium platform, the distilled NLLB-200 (600M) model, and Gemma-4-31B (Few-Shot | Direct). The annotator was presented with 100 round-robin matchups per translation direction and asked to indicate his preference between two anonymized system outputs, with the option of declaring a tie, and with the gold human reference injected as a hidden system. The resulting pairwise judgments were aggregated into a global ranking using the TrueSkill algorithm (Herbrich et al., 2007), with default priors ($\mu = 25.0$, $\sigma = 8.333$) and ties resolved via TrueSkill’s native draw mechanism. For each system, we report in Table 2 a conservative score ($\mu - 3\sigma$) and, for adjacent pairs in the ranking, the probability that the higher-ranked system’s true skill exceeds the lower-ranked system’s, computed in closed form from the difference of their posterior Gaussians.

These results show that human judgments closely mirror the trends observed in the automatic metrics, highlighting a stark asymmetry depending on the translation direction. In the FR → FRP di-

¹⁷To test whether these performance differences are statistically significant, we performed paired bootstrap resampling (Koehn, 2004). The Direct-over-Reasoning advantage is significant ($p < 0.05$) in ten of the twelve model/regime/direction comparisons. Two comparisons deviate: zero-shot Gemma-4-31B in FRP→FR, where the two modalities are statistically indistinguishable (+0.5 chrF++, n.s.), and few-shot Gemma-4-E2B in FRP→FR, where Reasoning significantly outperforms Direct (−2.0 chrF++, $p < 0.05$). Full pairwise significance matrices are given in Appendix C. We note that the significance matrices report *segment-level* mean chrF++ differences, whereas Table 1 reports corpus-level scores; the two diverge for the degenerate E2B Reasoning configurations, whose spurious chain-of-thought output is penalized far more heavily under corpus-level aggregation.

Direction	System	Cons. Score	μ	σ	Prob. > Next	Significance
FR → FRP	Hidden Human Reference	27.32	31.41	1.36	>99.9%	$p < 0.001$
	Apertium (Rule-Based)	20.53	23.94	1.14	67.3%	n.s.
	NLLB-200 Distilled (600M)	19.79	23.22	1.14	81.3%	n.s.
	Gemma-4-31B (Few-Shot)	18.29	21.77	1.16	—	—
FRP → FR	Hidden Human Reference	25.26	28.92	1.22	88.1%	n.s. ($p \geq 0.05$)
	Gemma-4-31B (Few-Shot)	23.51	26.94	1.14	97.5%	$p < 0.05$
	NLLB-200 Distilled (600M)	20.30	23.75	1.15	81.4%	n.s.
	Apertium (Rule-Based)	18.67	22.27	1.20	—	—

Table 2: TrueSkill human evaluation leaderboards for both translation directions. The Conservative Score represents the lower 99% credible bound of the system’s estimated true skill ($\mu - 3\sigma$). The “Prob. > Next” column indicates the probability of the system’s true skill being strictly greater than that of the system ranked immediately below it, alongside its corresponding statistical significance threshold (n.s. = not significant).

rection, the hidden human reference ($\mu = 31.41$) significantly outperformed all machine systems ($p < 0.001$). Among the automated approaches, the hand-crafted rules of Apertium ($\mu = 23.94$) and the Distilled NLLB model ($\mu = 23.22$) performed comparably, effectively tying for first place. Notably, the Distilled NLLB model outperformed Gemma-4-31B ($\mu = 21.77$). Conversely, in the FRP → FR direction, Gemma-4-31B Few-Shot Direct ($\mu = 26.94$) demonstrated overwhelming superiority over both NLLB and Apertium. Remarkably, the probability of superiority between the hidden human reference and Gemma-4-31B was only 88.14%, indicating that the LLM achieved near-human parity ($p \geq 0.05$) when decoding into standard French.

The automatic and human evaluations therefore converge, validating both the automatic metrics for this language pair and the quality of the reference translation, to an extent that is further discussed in the Limitations section. The resulting picture is asymmetric: translation into French is approaching human parity, while into Savoyard the human reference remains significantly preferred over all systems and the best baseline system is rule-based.

6 Related Work

The first MT evaluation dataset bearing the name “FLORES” was released in 2019 and targeted two resource-constrained Indo-Aryan languages: Nepali and Sinhala (Guzmán et al., 2019). Its source segments were extracted from Wikipedia articles, and professionally translated, with stringent iterative rounds of quality assurance. In order to mitigate the effects of translationese on MT evaluation (Zhang and Toral, 2019; Graham et al., 2020), source segments were also sampled from Nepali and Sinhala Wikipedia, rather than English alone.

The subsequent FLORES-101 dataset (Goyal et al., 2022) forms the actual basis of FLORES+. While it retained the core methodology of professional translation and rigorous quality assurance protocols, its 2,009 source segments bear no relation to those of the original FLORES, having been drawn instead from English-language Wikimedia projects (Wikitravel, Wikinews, Wikijunior) and translated into 101 target languages. This English-centric design made possible a fully aligned many-to-many evaluation across 10,100 language pairs, allowing researchers to assess regionally relevant, non-English translation directions.

FLORES-101 was further expanded into FLORES-200 in the context of the NLLB-200 project (NLLB Team, 2024), covering 200 languages as the name implies. The corpus, now curated by the Open Language Data Initiative (Burchell et al., 2024; Dale et al., 2025) under the name FLORES+, is regularly extended into new low-resource languages (Gordeev et al., 2024; Kuzhuget et al., 2024; Ali et al., 2024; Perez-Ortiz et al., 2024; Frontull et al., 2025; Lou et al., 2025; Samil and Adelani, 2026), with important contributions also including corrections to previously released partitions (Abdulmumin et al., 2024; Perez-Ortiz et al., 2024; Mæhlum et al., 2025; Oktem et al., 2025).

Other multi-parallel evaluation corpora include TICO-19 (Anastasopoulos et al., 2020), which targets the medical domain across 38 languages, and NTREX-128 (Federmann et al., 2022), which translates 1,997 WMT19 news segments (Barrault et al., 2019) into 128 languages.¹⁸ Similarly, WMT24++ (Deutsch et al., 2025) extends the WMT24 test set

¹⁸NTREX-128’s exposure to web crawlers on GitHub appears to have contaminated LLM pre-training data, limiting its use as a strict evaluation benchmark but making it a valuable few-shot training resource.

(Kocmi et al., 2024) to 55 languages. Despite potential GDPR constraints regarding its social media data, it uniquely provides both initial translations and final post-edits, making it ideal for studying translation dynamics (cf. Marmonier et al. 2026). Moving away from English-centric models, the BOUQuET dataset (Andrews et al., 2025) uses original texts handcrafted across eight pivot languages. Spanning 275 variants, it covers diverse registers—such as fiction and dialogues—while maintaining low structural complexity to encourage crowd-sourced extensions.

To the best of our knowledge, no machine translation evaluation corpora for Savoyard or the broader Arpitan dialect continuum currently exist, making the present dataset the first of its kind for this language family.

7 Conclusion

We have presented a Savoyard partition of the FLORES+ corpus, together with a range of baseline scores intended to give future work a point of comparison. Savoyard is but one variety within a much larger and diverse whole: the Arpitan dialect continuum extends across the French departments of eastern and central France (including varieties such as Bressan, Lyonnais, Forézien, and Dauphinois), French-speaking Switzerland (encompassing dialects like Vaudois, Genevois, Neuchâtelois, Fribourgeois, and Valaisan), and the Aosta Valley of Italy. All of these are low-resource varieties from the point of view of natural language processing, and their revitalization would likely benefit greatly from translation technology. We hope that the present work will stimulate similar efforts for these dialects elsewhere, contributing in turn to the preservation of humanity’s linguistic diversity.

Limitations

The main limitations of this work stem from the very scarcity of speakers that motivates it.

A single-translator corpus. The workload, initially to be shared between two translators, was ultimately borne almost entirely by one, the pool of expert translators being extremely small. This does not bear on the quality of the resulting corpus, as it was produced by a lexicographer of Savoyard working under the auspices of the *Institut de la langue savoyarde*, the body whose lexicographical and normative work defines the Median Savoyard

standard the partition targets. It does, however, bound what the corpus attests.

Human evaluation. The annotation campaign was conducted by the translator himself. The design was blind: outputs were anonymized and the annotator was not informed that the gold references had been injected as a hidden system. Some residual self-preference along stylistic lines cannot be excluded, since the norms against which the systems were judged are those the annotator/translator himself applies. The campaign is best read, accordingly, as establishing agreement with the ILS standard—which is the property our references are meant to have—rather than as an independent measure of acceptability across the entire Savoyard dialect continuum.

Confounds in the baselines. The translator contributed in the past to the French–Arpitan Apertium system. Although the devtest was translated from scratch, Apertium’s FR→FRP scores are plausibly inflated by this shared provenance.

Acknowledgments

This work was funded by the French *Agence Nationale de la Recherche* (ANR) under the project TraLaLaM (“ANR-23-IAS1-0006”), as well as by Inria under the “Défi”-type project COLaF. It was also partly funded by Benoît Sagot’s chair in the PRAIRIE-PSAI institute, funded by the ANR as part of the “France 2030” strategy under the reference ANR-23-IACL-0008.

This work was also granted access to the HPC resources of IDRIS under allocations 2025-AD011015117R2, 2025-AD011012254R5 and 2025-AD011017228 made by GENCI. The authors acknowledge the use of LLMs as coding assistants during the implementation of this work.

We are grateful to François Yvon for his help with the administrative aspects of this project. We also thank Franck Monod, Operations Manager at the *Institut de la langue savoyarde*, and Amandine Pinget, Data Scientist at the *Institut de la langue savoyarde*, for their assistance and support in coordinating this collaboration.

References

Idris Abdulmumin, Sthembiso Mkhwanazi, Mahlatse Mbooi, Shamsuddeen Hassan Muhammad, Ibrahim Said Ahmad, Neo Putini, Miehleketo Mathabela, Matimba Shingange, Tajuddeen Gwadabe,

- and Vukosi Marivate. 2024. [Correcting FLORES Evaluation Dataset for Four African Languages](#). In *Proceedings of the Ninth Conference on Machine Translation*, pages 570–578, Miami, Florida, USA. Association for Computational Linguistics.
- Felermينو Dario Mario Ali, Henrique Lopes Cardoso, and Rui Sousa-Silva. 2024. [Expanding FLORES+ Benchmark for More Low-Resource Settings: Portuguese-Emakhuwa Machine Translation Evaluation](#). In *Proceedings of the Ninth Conference on Machine Translation*, pages 579–592, Miami, Florida, USA. Association for Computational Linguistics.
- Dante Alighieri. 1996. *De vulgari eloquentia*. Number 5 in Cambridge Medieval Classics. Cambridge University Press, Cambridge.
- Hèctor Alòs. 2020. [GSOC 2020 rapport final: Traducteur bidirectionnel français-arpitan](#). Apertium Wiki. Accessed: 2026-08-03.
- Antonios Anastasopoulos, Alessandro Cattelan, Zi-Yi Dou, Marcello Federico, Christian Federmann, Dmitriy Genzel, Francisco Guzmán, Junjie Hu, Macduff Hughes, Philipp Koehn, Rosie Lazar, Will Lewis, Graham Neubig, Mengmeng Niu, Alp Öktem, Eric Paquin, Grace Tang, and Sylwia Tur. 2020. [TICO-19: the Translation Initiative for COvid-19](#). In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*, Online. Association for Computational Linguistics.
- Pierre Andrews, Mikel Artetxe, Mariano Coria Meglioli, Marta R. Costa-jussà, Joe Chuang, David Dale, Mark Duppenthaler, Nathaniel Paul Ekberg, Cynthia Gao, Daniel Edward Licht, Jean Maillard, Alexandre Mourachko, Christophe Ropers, Safiyyah Saleem, Eduardo Sánchez, Ioannis Tsiamas, Arina Turkatenco, Albert Ventayol-Boada, and Shireen Yates. 2025. [BOUQuET : Dataset, Benchmark and Open initiative for Universal Quality Evaluation in Translation](#). In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 27515–27535, Suzhou, China. Association for Computational Linguistics.
- Graziadio Isaia Ascoli. 1878. Schizzi franco-provenzali. *Archivio Glottologico Italiano*, 3:61–120.
- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. Neural Machine Translation by Jointly Learning to Align and Translate. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR)*.
- Loïc Barrault, Ondřej Bojar, Marta R. Costa-jussà, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Philipp Koehn, Shervin Malmasi, Christof Monz, Mathias Müller, Santanu Pal, Matt Post, and Marcos Zampieri. 2019. [Findings of the 2019 Conference on Machine Translation \(WMT19\)](#). In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 1–61, Florence, Italy. Association for Computational Linguistics.
- Michel Bert and James Costa. 2009. *Etude fora : Francoprovençal et occitan en rhône-alpes*. Rapport de recherche, L’Institut Pierre Gardette.
- Natalia Bichurina. 2019. *L’émergence du franco-provençal : langue minoritaire et communauté autour du Mont-Blanc*. Maison des sciences de l’homme d’Aquitaine. Préface de Manuel Meune.
- Natalia Bichurina and Christiane Dunoyer. 2021. [Le francoprovençal en savoie : Histoire et pratiques contemporaines](#). *Les Dossiers du Musée Savoisien : Revue numérique*, 7. [En ligne].
- L. Boutillier. 2022. [Le terme savoyard « monchu » intègre le dictionnaire](#). *La Savoie*.
- Peter F. Brown, John Cocke, Stephen A. Della Pietra, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer, and Paul S. Roossin. 1990. A Statistical Approach to Machine Translation. *Comput. Linguist.*, 16(2):79–85.
- Laurie Burchell, Jean Maillard, Antonios Anastasopoulos, Christian Federmann, Philipp Koehn, and Skyler Wang. 2024. [Findings of the WMT 2024 Shared Task of the Open Language Data Initiative](#). In *Proceedings of the Ninth Conference on Machine Translation*, pages 110–117, Miami, Florida, USA. Association for Computational Linguistics.
- Isaac Caswell, Elizabeth Nielsen, Jiaming Luo, Colin Cherry, Geza Kovacs, Hadar Shemtov, Partha Talukdar, Dinesh Tewari, Baba Mamadi Diane, Djib-rila Diane, Solo Farabado Cissé, Koulako Moussa Doumbouya, Edoardo Ferrante, Alessandro Guasoni, Christopher Homan, Mamadou K. Keita, Sudhamoy DebBarma, Ali Kuzhuget, David Anugraha, and 5 others. 2025. [SMOL: Professionally Translated Parallel Data for 115 Under-represented Languages](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 1103–1123, Suzhou, China. Association for Computational Linguistics.
- David Dale, Laurie Burchell, Jean Maillard, Idris Abdumumin, Antonios Anastasopoulos, Isaac Caswell, and Philipp Koehn. 2025. [Findings of the WMT 2025 Shared Task of the Open Language Data Initiative](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 495–502, Suzhou, China. Association for Computational Linguistics.
- Joseph Désormaux. 1981. *Bibliographie méthodique des parlers de Savoie: langue et littérature*. Laffitte, Marseille, France. Facsimile reprint of the original 1922 edition published in Annecy by Imprimerie commerciale.
- Daniel Deutsch, Eleftheria Briakou, Isaac Rayburn Caswell, Mara Finkelstein, Rebecca Galor, Juraj Juraska, Geza Kovacs, Alison Lui, Ricardo Rei, Jason Riesa, Shruti Rijhwani, Parker Riley, Elizabeth Salesky, Firas Trabelsi, Stephanie Winkler, Biao Zhang, and Markus Freitag. 2025. [WMT24++: Expanding the Language Coverage of WMT24 to 55](#)

- [Languages & Dialects](#). In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 12257–12284, Vienna, Austria. Association for Computational Linguistics.
- Christiane Dunoyer, editor. 2019. *Regards croisés sur la standardisation du francoprovençal: Actes de la conférence annuelle sur l'activité scientifique du Centre d'Études francoprovençales (Saint-Nicolas, 11 novembre 2017)*. Région autonome Vallée d'Aoste, Aoste.
- Alain Favre. 2026. *Diccionèro de fata: Savoyârd – Francès / Français – Savoyard*. Yoran Embanner, Spézet.
- Alain Favre and Dominique Stich. 2005. *Diccionèro de fata / Minidico: Savoyârd / Francès – Français / Savoyard*. Yoran Embanner, Spézet.
- Christian Federmann, Tom Kocmi, and Ying Xin. 2022. [NTREX-128 – News Test References for MT Evaluation of 128 Languages](#). In *Proceedings of the First Workshop on Scaling Up Multilingual Evaluation*, pages 21–24, Online. Association for Computational Linguistics.
- Mikel L. Forcada and Francis M. Tyers. 2016. [Aperium: a free/open source platform for machine translation and basic language technology](#). In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation: Projects/Products*, Riga, Latvia. Baltic Journal of Modern Computing.
- Samuel Frontull, Thomas Ströhle, Carlo Zoli, Werner Pescosta, Ulrike Frenademez, Matteo Ruggeri, Daria Valentin, Karin Comploj, Gabriel Perathoner, Silvia Liotto, and Paolo Anvidalfarei. 2025. [Bringing Ladin to FLORES+](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 1061–1071, Suzhou, China. Association for Computational Linguistics.
- Google DeepMind. 2026. Gemma 4 model card. https://ai.google.dev/gemma/docs/core/model_card_4.
- Isai Gordeev, Sergey Kuldin, and David Dale. 2024. [FLORES+ Translation and Machine Translation Evaluation for the Erzya Language](#). In *Proceedings of the Ninth Conference on Machine Translation*, pages 614–623, Miami, Florida, USA. Association for Computational Linguistics.
- Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc’Aurelio Ranzato, Francisco Guzmán, and Angela Fan. 2022. [The Flores-101 Evaluation Benchmark for Low-Resource and Multilingual Machine Translation](#). *Transactions of the Association for Computational Linguistics*, 10:522–538.
- Yvette Graham, Barry Haddow, and Philipp Koehn. 2020. [Statistical Power and Translationese in Machine Translation Evaluation](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 72–81, Online. Association for Computational Linguistics.
- Henri Grégoire. 1794. [Rapport sur la nécessité et les moyens d’anéantir les patois et d’universaliser l’usage de la langue française](#). Historical document, Convention nationale. Available on Wikisource.
- Francisco Guzmán, Peng-Jen Chen, Myle Ott, Juan Pino, Guillaume Lample, Philipp Koehn, Vishrav Chaudhary, and Marc’Aurelio Ranzato. 2019. [The FLORES Evaluation Datasets for Low-Resource Machine Translation: Nepali–English and Sinhala–English](#). In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6098–6111, Hong Kong, China. Association for Computational Linguistics.
- Barry Haddow, Rachel Bawden, Antonio Valerio Miceli Barone, Jindřich Helcl, and Alexandra Birch. 2022. [Survey of Low-Resource Machine Translation](#). *Computational Linguistics*, 48(3):673–732.
- Ralf Herbrich, Tom Minka, and Thore Graepel. 2007. [Trueskill™: A bayesian skill rating system](#). In Bernhard Schölkopf, John Platt, and Thomas Hofmann, editors, *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*. The MIT Press.
- Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. 2020. [The State and Fate of Linguistic Diversity and Inclusion in the NLP World](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6282–6293, Online. Association for Computational Linguistics.
- Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, Barry Haddow, Marzena Karpinska, Philipp Koehn, Benjamin Marie, Christof Monz, Kenton Murray, Masaaki Nagata, Martin Popel, Maja Popović, and 3 others. 2024. [Findings of the WMT24 General Machine Translation Shared Task: The LLM Era Is Here but MT Is Not Solved Yet](#). In *Proceedings of the Ninth Conference on Machine Translation*, pages 1–46, Miami, Florida, USA. Association for Computational Linguistics.
- Philipp Koehn. 2004. [Statistical Significance Tests for Machine Translation Evaluation](#). In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 388–395, Barcelona, Spain. Association for Computational Linguistics.
- Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. [Statistical phrase-based translation](#). In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL ’03, page 48–54, USA. Association for Computational Linguistics.

- Andres Kristol. 2016. [Francoprovençal](#). In Adam Ledgeway and Martin Maiden, editors, *The Oxford Guide to the Romance Languages*. Oxford University Press.
- Ali Kuzhuget, Airana Mongush, and Nachyn-Enkhedorzhu Oorzhak. 2024. [Enhancing Tuvan Language Resources through the FLORES Dataset](#). In *Proceedings of the Ninth Conference on Machine Translation*, pages 593–599, Miami, Florida, USA. Association for Computational Linguistics.
- Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph E. Gonzalez, Hao Zhang, and Ion Stoica. 2023. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the ACM SIGOPS 29th Symposium on Operating Systems Principles*.
- Mark Liberman. 2010. [Obituary: Fred Jelinek](#). *Computational Linguistics*, 36(4):595–599.
- Andrés Lou, Juan Antonio Pérez-Ortiz, Felipe Sánchez-Martínez, Miquel Esplà-Gomis, and Víctor M. Sánchez-Cartagena. 2025. [FLORES+ Mayas: Generating Textual Resources to Foster the Development of Language Technologies for Mayan Languages](#). In *Proceedings of Machine Translation Summit XX: Volume 2*, pages 89–90, Geneva, Switzerland. European Association for Machine Translation.
- Petter Mæhlum, Anders Næss Evensen, and Yves Scherrer. 2025. [Improved Norwegian Bokmål Translations for FLORES](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 1124–1132, Suzhou, China. Association for Computational Linguistics.
- Jean Maillard, Cynthia Gao, Elahe Kalbassi, Kaushik Ram Sadagopan, Vedanuj Goswami, Philipp Koehn, Angela Fan, and Francisco Guzman. 2023. [Small Data, Big Impact: Leveraging Minimal Data for Effective Machine Translation](#). In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2740–2756, Toronto, Canada. Association for Computational Linguistics.
- Malik Marmonier, Rachel Bawden, and Benoît Sagot. 2026. [Testing the Deliteralization Hypothesis in Human and Machine Translation](#). *Preprint*, arXiv:2605.25686.
- Malik Marmonier, Benoît Sagot, and Rachel Bawden. 2025. [A French Version of the OLDI Seed Corpus](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 1048–1060, Suzhou, China. Association for Computational Linguistics.
- Ministère de l’Éducation nationale, de la Jeunesse et des Sports. 2021. Bulletin officiel n° 47 du 16 décembre 2021. https://www.education.gouv.fr/sites/default/files/document/BOENJS_47_1423493.pdf-311421.pdf. Accessed: July 9, 2026.
- NLLB Team. 2024. [Scaling neural machine translation to 200 languages](#). *Nature*, 630(8018):841–846.
- Alp Oktem, Mohamed Aymane Farhi, Brahim Essaidi, Naceur Jabouja, and Farida Boudichat. 2025. [Correcting the Tamazight Portions of FLORES+ and OLDI Seed Datasets](#). In *Proceedings of the Tenth Conference on Machine Translation*, pages 1072–1080, Suzhou, China. Association for Computational Linguistics.
- Michèle Oliuéri and Patrick Sauzet. 2016. [Southern Gallo-Romance \(Occitan\)](#). In Adam Ledgeway and Martin Maiden, editors, *The Oxford Guide to the Romance Languages*. Oxford University Press.
- Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. [Bleu: a Method for Automatic Evaluation of Machine Translation](#). In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.
- Juan Antonio Perez-Ortiz, Felipe Sánchez-Martínez, Víctor M. Sánchez-Cartagena, Miquel Esplà-Gomis, Aaron Galiano Jimenez, Antoni Oliver, Claudi Aventín-Boya, Alejandro Pardos, Cristina Valdés, Jusèp Loís Sans Socasau, and Juan Pablo Martínez. 2024. [Expanding the FLORES+ Multilingual Benchmark with Translations for Aragonese, Aranese, Asturian, and Valencian](#). In *Proceedings of the Ninth Conference on Machine Translation*, pages 547–555, Miami, Florida, USA. Association for Computational Linguistics.
- Maja Popović. 2015. [chrF: character n-gram F-score for automatic MT evaluation](#). In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.
- Maja Popović. 2017. [chrF++: words helping character n-grams](#). In *Proceedings of the Second Conference on Machine Translation*, pages 612–618, Copenhagen, Denmark. Association for Computational Linguistics.
- Matt Post. 2018. [A Call for Clarity in Reporting BLEU Scores](#). In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Belgium, Brussels. Association for Computational Linguistics.
- République française. 1951. Loi n° 51-46 du 11 janvier 1951 relative à l’enseignement des langues et dialectes locaux. <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886638>. Journal officiel de la République française, 13 janvier 1951, p. 483. Accessed: 2026-06-05.
- Hadia Mohmmadosman Ahmed Samil and David Ifeoluwa Adelani. 2026. [Sudanese-Flores: Extending FLORES+ to Sudanese Arabic Dialect](#). In *Proceedings of the 7th Workshop on African Natural Language Processing (AfricaNLP 2026)*, pages 243–247,

- Rabat, Morocco. Association for Computational Linguistics.
- Tove Skutnabb-Kangas. 2000. *Linguistic Genocide in Education—Or Worldwide Diversity and Human Rights?* Lawrence Erlbaum Associates, Mahwah, NJ. Includes bibliographical references and index.
- John Charles Smith. 2016. [French and northern Gallo-Romance](#). In Adam Ledgeway and Martin Maiden, editors, *The Oxford Guide to the Romance Languages*. Oxford University Press.
- Matthew Snover, Bonnie Dorr, Rich Schwartz, Linnea Micciulla, and John Makhoul. 2006. [A Study of Translation Edit Rate with Targeted Human Annotation](#). In *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas: Technical Papers*, pages 223–231, Cambridge, Massachusetts, USA. Association for Machine Translation in the Americas.
- Dominique Stich. 2001. *Francoprovençal : proposition d'une orthographe supra-dialectale standardisée*. Thèse de doctorat, Université Paris 5. Moteur de recherche des thèses françaises: theses.fr.
- Dominique Stich. 2003. *Dictionnaire francoprovençal / français, français / francoprovençal*. Éditions Le Carré, Thonon-les-Bains.
- Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'14, page 3104–3112, Cambridge, MA, USA. MIT Press.
- Gaston Tuaillon. 1983. *Le Francoprovençal: Progrès d'une définition*. Centre d'Études Francoprovençales « René Willien », Saint-Nicolas (Aoste).
- Gaston Tuaillon. 1984. La graphie de Conflans pour le patois savoyard. *Les Nouvelles du Centre d'Études francoprovençales « René Willien »*, page 5.
- Gaston Tuaillon. 1989. Le francoprovençal, langue oubliée. In Geneviève Vernes, editor, *Vingt-cinq communautés linguistiques de la France*, volume 1, page 204. L'Harmattan, Paris.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NeurIPS'17, page 6000–6010, Red Hook, NY, USA. Curran Associates Inc.
- Roger Viret. 2015. *Graméra savoyârda: Arbané*. Institut de la langue savoyarde, Challes-les-Eaux, France.
- Roger Viret. 2019. *Dictionnaire français–savoyard*. Troisième édition revue et augmentée.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, and 3 others. 2020. [Transformers: State-of-the-Art Natural Language Processing](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online. Association for Computational Linguistics.
- Mike Zhang and Antonio Toral. 2019. [The Effect of Translationese in Machine Translation Test Sets](#). In *Proceedings of the Fourth Conference on Machine Translation (Volume 1: Research Papers)*, pages 73–81, Florence, Italy. Association for Computational Linguistics.

A Data Distillation prompt

This appendix provides an example of the template used for the *Distilled (Gold+Silver)* dataset generation, where Gemma-4-31B was provided with French–Arpitan–Savoyard few-shot triplet demonstrations to post-edit the rule-based Apertium drafts.

Gemma-4-31B Distillation Prompt Template

Savoyard (endonym: *savoyârd*) is a Franco-Provençal dialect of the Gallo-Romance family. It is spoken in some territories of the historical Duchy of Savoy, nowadays a geographic area spanning Savoie and Haute-Savoie, France, and the Canton of Geneva, Switzerland.

You are to translate a French segment of text into Savoyard. To assist you in this task, we will provide you with translation hypotheses produced by Apertium, a rule-based machine translation platform. It will provide a rough candidate translation from French into an Arpitan macrolanguage. Your task is to post-edit and refine this candidate to produce a fluent, orthographically correct, and natural Savoyard translation.

To guide your style and vocabulary, here is a parallel corpus of expert translations demonstrating the task at hand:

<parallel_corpus>

French: *Des scientifiques de l'école de*

médecine de l'université de Stanford ont annoncé ce lundi la création d'un nouvel outil de diagnostic, qui permettrait de différencier les cellules en fonction de leur type. Il s'agit d'une petite puce imprimable, qui peut être produite au moyen d'une imprimante à jet d'encre standard, pour un coût d'environ un cent de dollar pièce.

Apertium Candidate: Des scientificos de l'ècoula de mèdecina de l'univèrsitàt de Stanford ant anonciè ceti delon la crèacion d'un novèl outil de diagnostic, que pèmetrèt de difèrenciér les cèlules en fonccion de lor tipo. Il s'agèt d'una petit puge imprimable, que pôt être fêta u moyen d'una èstampiosa a jèt d'encro estandârd, por un côte d'enveron un cent de dolâr piéce.

Savoyard: Des scientificos de l'ècoula de mèdecina de l'univèrsitàt de Stanford ant anonciè ceti delon la crèacion d'un novèl outil de diagnostic, que pèmetrèt de difèrenciér les cèlules en fonccion de lor tipo. O est una petiôta puge emprimâbla, que pôt être fêta avouéc una èstampiosa a jèt d'encro estandârd, por un côte d'enveron un cent de dolâr piéce.

...

French: Les compagnies aériennes qui les proposent sont notamment Air Canada, Delta Air Lines, Lufthansa pour les vols en provenance des États-Unis ou du Canada, et WestJet.

Apertium Candidate: Les compagnies d'aviacion que los propôsont sont spécialement Air Canadâ, Dèlta Air Lines, Lufthansa por los vòls en provegnance des Ètats-Unis ou du Canadâ, et WestJet.

Savoyard: Les compagnies d'aviacion que los propôsont sont spécialement Air Canada, Dèlta Air Lines, Lufthansa por los vòls que vegnont des Ètats-Unis ou du Canadâ, et WestJet.

French: Dans tous les cas, vous devez réserver par téléphone directement

auprès de la compagnie aérienne.

Apertium Candidate: Dens tôs los câs, vos dête resèrvâr per tèlèfono dirèctament uprès de la compagnie d'aviacion.

Savoyard: Diens tôs los câs, vos dêde resèrvâr per tèlèfono dirèctament uprès de la compagnie d'aviacion.

</parallel_corpus>

You must only output your final translation between the XML tags and absolutely nothing else. Do not include any reasoning, notes, or introductions. The expected XML format is: <translation>your translation here</translation>

Here is the segment you must translate:

French: La première avancée majeure en termes d'abstraction fut l'utilisation de chiffres pour représenter les nombres.

Apertium Candidate: La première avancîe majore en tèrmenos d'abstraccion at étâ l'usâjo de chifros por représentar los nombros.

Savoyard:

B LLM Translation Prompt Templates

This appendix contains the prompts used for the evaluation of the Gemma-4 models (31B, E4B, and E2B) described in Section 4. We provide examples in the FR→FRP translation direction; the FRP→FR differs minimally, for instance in the order in which dev set bitexts are presented.

B.1 Zero-Shot Templates

B.1.1 Direct (No-Think) Modality

The following template was used for the zero-shot direct generation condition. The model was instructed to output the final translation immediately within XML tags.

Zero-Shot Direct Translation Prompt

Savoyard (endonym: *savoyârd*) is a Franco-Provençal dialect of the Gallo-Romance family. It is spoken in some territories of the historical Duchy of Savoy, nowadays a geographic area spanning Savoie and

Haute-Savoie, France and the Canton of Geneva, Switzerland.

You are to translate the following French text segment into Savoyard. **You must only output the translation between the XML tags and absolutely nothing else. Do not include any reasoning, notes, or introductions.** The expected XML format is: <translation>your translation here</translation>

Segment to translate:

Les travailleurs doivent souvent obtenir l'accord de leurs supérieurs pour toute décision qu'ils prennent et sont censés obéir aux instructions de leurs supérieurs sans poser de questions.

B.1.2 Reasoning Modality

The following template was used for the zero-shot reasoning condition, explicitly allowing the model to utilize an internal monologue prior to emitting the translation.

Zero-Shot Reasoning Translation Prompt

Savoyard (endonym: *savoyârd*) is a Franco-Provençal dialect of the Gallo-Romance family. It is spoken in some territories of the historical Duchy of Savoy, nowadays a geographic area spanning Savoie and Haute-Savoie, France and the Canton of Geneva, Switzerland.

You are to translate the following French text segment into Savoyard. **You can reflect on the task at hand prior to producing your final translation.** You must output your final translation between XML tags thusly: <translation>your translation here</translation>

Segment to translate:

Les travailleurs doivent souvent obtenir l'accord de leurs supérieurs pour toute décision qu'ils prennent et sont censés obéir aux instructions de leurs supérieurs sans poser de questions.

B.2 Few-Shot Templates

B.2.1 Direct (No-Think) Modality

For the few-shot direct condition, the prompt was appended with the entire structured *dev* partition bitext as in-context demonstrations, immediately followed by the target source query.

Few-Shot Direct Translation Prompt

Savoyard (endonym: *savoyârd*) is a Franco-Provençal dialect of the Gallo-Romance family. It is spoken in some territories of the historical Duchy of Savoy, nowadays a geographic area spanning Savoie and Haute-Savoie, France and the Canton of Geneva, Switzerland.

You will be asked to translate a French text segment into Savoyard. **You must only output the translation between the XML tags and absolutely nothing else. Do not include any reasoning, notes, or introductions.** The expected XML format is: <translation>your translation here</translation>

To assist you in this task, here is a small parallel corpus. It should dispel doubts you might have regarding the orthography and grammar of this romance dialect:

<parallel_corpus>

French: *Des scientifiques de l'école de médecine de l'université de Stanford ont annoncé ce lundi la création d'un nouvel outil de diagnostic, qui permettrait de différencier les cellules en fonction de leur type. Il s'agit d'une petite puce imprimable, qui peut être produite au moyen d'une imprimante à jet d'encre standard, pour un coût d'environ un cent de dollar pièce.*

Savoyard: *Des scientifiques de l'ècoula de mèdecina de l'univèrsitàt de Stanford ant anonciè ceti delon la crèacion d'un novél outil de diagnostico, que pèmetrèt de difèrenciér les cèlules en fonccion de lor tipo. O est una petiôta puge emprimâbla, que pôèt ètre fêta avouéc una èstampiosa a jèt d'encro estandârd, por un côte d'enveron un cent de dolâr piéce.*

French: *Selon les chercheurs principaux, cela pourrait permettre une détection précoce du cancer, de la tuberculose, du VIH et du paludisme chez les patients des pays à faible revenu, où le taux de survie à des maladies telles que le cancer du sein peut être deux fois moins élevé que dans les pays riches.*

Savoyard: *D'après los chërchiors principaux, cen porrêt pèrmetre una dètèccion couètosa du chancro, de la tubèrculosa, du VIH et du paludismo chiéz los malâdos des pays a fêbla renta, yô que lo tôx de survia a des maladies coment lo chancro du nènè pôût être doux côps muens hiôt que diens los pays rechos.*

...

French: *Dans tous les cas, vous devez réserver par téléphone directement auprès de la compagnie aérienne.*

Savoyard: *Diens tôs los câs, vos dêde resèrvan per tèlèfono dirèctament uprès de la compagnie d'aviacion.*
</parallel_corpus>

Here is the French segment you must translate into **Savoyard**:

Les travailleurs doivent souvent obtenir l'accord de leurs supérieurs pour toute décision qu'ils prennent et sont censés obéir aux instructions de leurs supérieurs sans poser de questions.

A reminder that you must only output the translation between the XML tags and absolutely nothing else. Do not include any reasoning, notes, or introductions. Format: <translation>your translation here</translation>

B.2.2 Reasoning Modality

For the few-shot reasoning condition, the system prompt instructed the model to reason before outputting the final tags, combined with the structural sequence of the *dev* bitext.

Few-Shot Reasoning Translation Prompt

Savoyard (endonym: *savoyârd*) is a Franco-Provençal dialect of the Gallo-Romance family. It is spoken in some territories of the historical Duchy of Savoy, nowadays a geographic area spanning Savoie and Haute-Savoie, France and the Canton of Geneva, Switzerland.

You will be asked to translate a French text segment into Savoyard. **You can reflect on the task at hand prior to producing your final translation.** You must output your final translation between XML tags thusly: <translation>your translation here</translation>

To assist you in this task, here is a small parallel corpus. It should dispel doubts you might have regarding the orthography and grammar of this romance dialect:

<parallel_corpus>

French: *Des scientifiques de l'école de médecine de l'université de Stanford ont annoncé ce lundi la création d'un nouvel outil de diagnostic, qui permettrait de différencier les cellules en fonction de leur type. Il s'agit d'une petite puce imprimable, qui peut être produite au moyen d'une imprimante à jet d'encre standard, pour un coût d'environ un cent de dollar pièce.*

Savoyard: *Des scientifiques de l'ècoula de mèdecina de l'univèrsitàt de Stanford ant anonciê ceti delon la crèacion d'un novél outil de diagnostico, que pèrmetrèt de difèrenciér les cèlules en fonccion de lor tipo. O est una petiôta puge emprimâbla, que pôût être fêta avouéc una èstampiosa a jèt d'encro estandârd, por un côte d'enveron un cent de dolâr pièce.*

French: *Selon les chercheurs principaux, cela pourrait permettre une détection précoce du cancer, de la tuberculose, du VIH et du paludisme chez les patients des pays à faible revenu, où le taux de survie à des maladies telles que le cancer du sein peut être deux fois moins élevé que dans les*

pays riches.

Savoyard: *D'après los chèrchiors principâls, cen porrêt pèrmètre una dètèccion couètosa du chancro, de la tubèrculosa, du VIH et du paludismo chiéz los malâdos des payis a fêbla renta, yô que lo tôx de survia a des maladies coment lo chancro du nè nè pôt être doux côps muens hiôt que diens los payis rechos.*

...

French: *Dans tous les cas, vous devez réserver par téléphone directement auprès de la compagnie aérienne.*

Savoyard: *Diens tôs los câs, vos dêde resèrvar per tèlèfono dirèctament uprés de la compagnie d'aviacion.*

</parallel_corpus>

Here is the French segment you must translate into **Savoyard**:

Les travailleurs doivent souvent obtenir l'accord de leurs supérieurs pour toute décision qu'ils prennent et sont censés obéir aux instructions de leurs supérieurs sans poser de questions.

A reminder that you can reflect on the task at hand prior to producing your final translation but must output your final translation between XML tags thusly: <translation>your translation here</translation>

C Statistical Significance Matrices

To assess whether the observed performance margins between systems are statistically meaningful, we conduct paired bootstrap resampling with 1,000 iterations. Figures 4 and 5 illustrate the mean chrF++ differences between all evaluated configurations for both translation directions.

Cells marked with an asterisk (*) indicate that the performance delta between the row system and column system is statistically significant at $p < 0.05$.

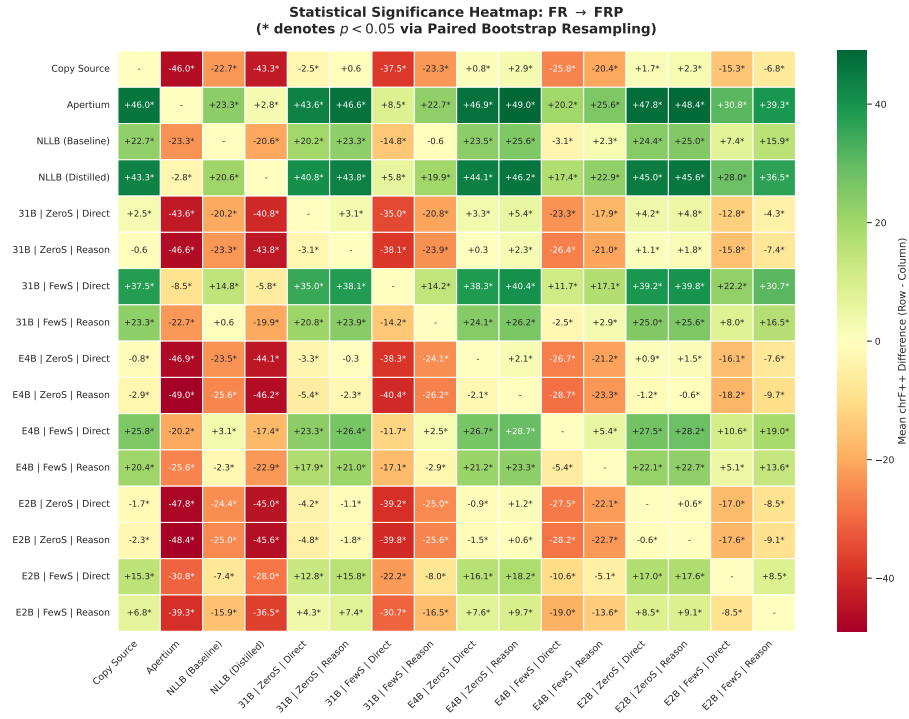


Figure 4: Statistical Significance Heatmap: FR → FRP

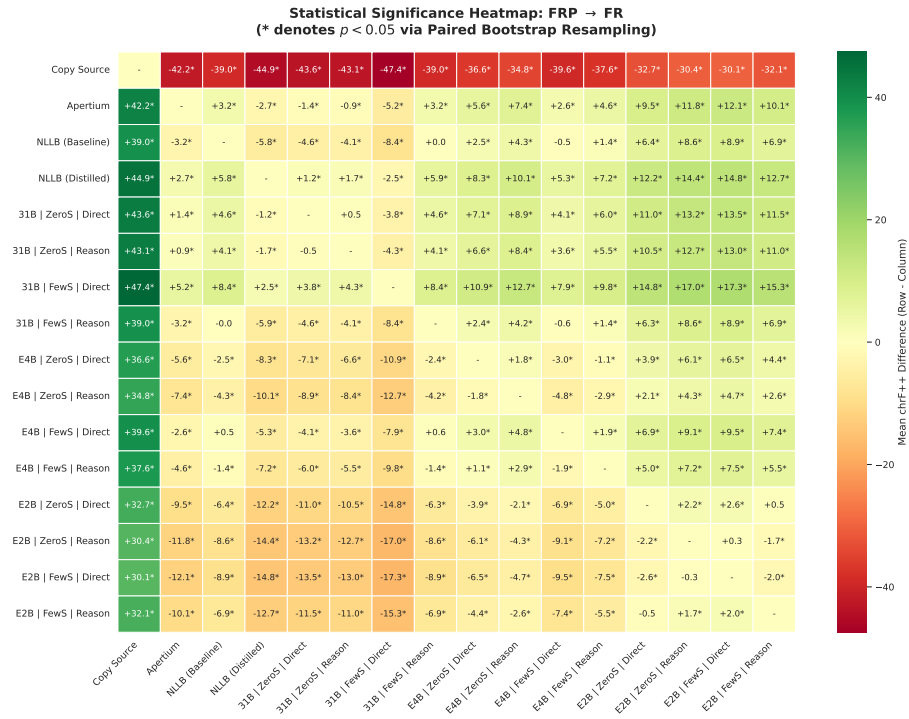


Figure 5: Statistical Significance Heatmap: FRP → FR

D Data Sample

We provide a random sample of the compiled dataset in Table 3.

ID	English Source	French Translation	Savoyard Translation
992	The tourist season for the hill stations generally peaks during the Indian summer.	<i>La saison touristique des stations de montagne culmine généralement pendant l'été indien.</i>	<i>La sêson touristica des stacions de montagne avente son maximon en gènèral pendent lo chòd-temps indien.</i>
553	Middle distance running is a relatively inexpensive sport; however, there are many misconceptions regarding the few pieces of equipment required to participate.	<i>Le demi-fond est un sport relativement peu coûteux. L'idée qu'il nécessiterait un équipement basique est aussi fausse que la plupart des mythes qui entourent ce sport.</i>	<i>Lo demi-fond est un sport .que côte pas tant. L'idè qu'il arèt fôta d'una forniture basica est asse fôssa coment la plepârt des mitos qu'entouront ceti sport.</i>
152	Virgin have only purchased the 'good bank' of Northern Rock, not the asset management company.	<i>Virgin n'a acheté que la «banque saine» de Northern Rock, pas la société de gestion d'actifs.</i>	<i>Virgin at achetâ solament la «banca sana» de Northern Rock, pas la sociêtât de maneyence d'actifs.</i>
189	The typhoon is estimated to be moving toward China at eleven kph.	<i>On estime que le typhon se dirige vers la Chine à une vitesse de 11 km/h.</i>	<i>On èstime que lo tifon s'aduit vers la China a una vitèsse de 11 km/h.</i>
463	Unfortunately, when you make a DVD, it's borders will most likely be cut too, and if the video had subtitles too close to the bottom, they won't be fully shown.	<i>Malheureusement, lorsque vous réalisez un DVD, ses bords seront probablement coupés aussi, et si la vidéo comportait des sous-titres trop proches du bas de l'écran, ils n'apparaîtront pas.</i>	<i>Mâlherosament, quand vos fabrecâd un DVD, sos bôrds seront probâblement copàs asse-ben, et se la vidèò at des sot-titros trop prôches du bâs de l'ècran, ils sè vèrront pas.</i>
953	Airlines such as Emirates, Etihad Airways, Qatar Airways & Turkish Airlines have greatly expanded their services to Africa, and offer connections to many major African cities at competitive rates than other European airlines.	<i>Les compagnies aériennes telles que Airlines such as Emirates, Etihad Airways, Qatar Airways et Turkish Airlines ont grandement développé leur offres en Afrique, et offrent des liaisons vers les villes africaines majeures à des tarifs compétitifs comparés aux autres compagnies aériennes européennes.</i>	<i>Les compagnies d'aviacion coment Emirates, Etihad Airways, Qatar Airways et Turkish Airlines ant rudament dèvelopâ lor ofres en Africa, et ofront des lièsons vers les veles afriquènes mès grantes a des prix compétitifs comparâ ux compagnies d'aviacion eropèènes.</i>
48	Mr Reid managed to drive the New Zealand's A1GP car, Black Beauty at speeds over 160km/h seven times over the bridge.	<i>M. Reid est parvenu au volant de la voiture néo-zélandaise de l'A1 GP dénommée Black Beauty, à dépasser à 7 reprises une vitesse de 160 km/h au dessus du pont.</i>	<i>M. Reid est aventâ u volant de la vouètura nèò-zèlandèsa de l'A1 GP apelâye Black Beauty, a dèpassar a 7 reprèses una vitèsse de 160 km/h u dessus du pont.</i>
406	Yes! King Tutankhamun, sometimes referred to as "King Tut" or "The Boy King", is one of the most well known ancient Egyptian kings in modern times.	<i>Oui! Le roi Toutankhamon, parfois appelé «Roi Tout» ou «L'Enfant roi», est l'un des anciens monarques égyptiens les plus connus des temps modernes.</i>	<i>Ouè! Lo rè Toutânkhamon, des côps apelâ «Rê Tot» ou «L'Enfant rê», est yon des ancians monarcos ègipcians mès famox des temps modèrns.</i>
937	In some federal countries, such as the United States and Canada, income tax is levied both at the federal level and at the local level, so the rates and brackets can vary from region to region.	<i>Dans certains pays fédéraux tels que les États-Unis et le Canada, l'impôt sur le revenu est prélevé à la fois au niveau fédéral et au niveau local, de sorte que les taux et les tranches d'imposition peuvent varier d'une région à l'autre.</i>	<i>Diens des payis fèdèrâls coment los États-Unis et lo Canadâ, l'emposicion sur la renta est prèlevâye a côp u nivèl fèdèrâl et u nivèl local, de façon que los tôx et les trenches d'emposicion pouessont vareyer d'una règion a l'ôtro.</i>
86	The amount and thickness of the pack ice, according to Pittman, is the worst it has been for sealers in the past 15 years.	<i>Selon M. Pittman, la quantité et l'épaisseur de la banquise sont les pires qu'aient connues les chasseurs de phoques au cours des 15 dernières années.</i>	<i>D'après M. Pittman, la quantitât et l'èpèssior de la banquisa sont les pires qu'èyont cognèssues los chasciors de foques diens los 15 dèrrièrs ans.</i>

Table 3: Random sample of ten sentence pairs from the *dev* partition of our Savoyard version of the FLORES+ corpus.