

Linguistique comparée des langues modernes

L4LM42 LF, jeudi 16h30-18h30, Malesherbes

CM : André THIBAUT

TD : Marc DUVAL

Première semaine : Séance d'introduction. Présentation du plan de cours. La famille des langues indoeuropéennes : aperçu général.

Comme l'intitulé de ce cours l'annonce, nous allons parler ce semestre-ci de linguistique en général, et de certaines langues en particulier ; cela dit, rassurez-vous, vous n'avez pas besoin de connaître au préalable toutes les langues dont nous allons parler pour pouvoir suivre le cours. Nous les décrirons de l'extérieur, comme l'entomologiste qui dissèque un insecte sous la loupe pour en découvrir et en décrire le plus précisément possible la structure. De toute façon, nous aurons toujours d'abord recours à des exemples tirés de la langue française pour illustrer les phénomènes au programme. En outre, un coup d'oeil au plan de cours vous montrera que l'on va consacrer plusieurs séances à des questions de terminologie et, par conséquent, de concepts : il ne s'agira pas, dans un premier temps, de décortiquer des langues en particulier, mais bien de se mettre d'accord sur ce que l'on entend par "langue", un des mots les plus polysémiques qui soient, même – et surtout ! – pour les linguistes. Ce n'est que dans une seconde étape que nous aborderons ce que l'on appelle la typologie linguistique, c'est-à-dire une activité de classification des langues du monde selon un ensemble de critères formels et fonctionnels.

Avant même de traiter ces questions, nous allons d'abord jeter un coup d'oeil à un premier type de typologie, de nature **génétique**. On peut en effet distinguer deux façons complètement différentes de classer les langues du monde. Il est possible – c'est ce qu'on fera dans la seconde moitié du semestre – de faire l'inventaire des différences et des ressemblances entre les différentes langues du monde, sans se demander si elles sont dues au hasard, d'une part, ou d'autre part à une origine commune ou à des phénomènes de contacts de langues. Mais si l'on se libère de cette étroite vision synchronique pour adopter une vue cavalière, panoramique, panchronique des phénomènes structurels, on voit qu'au-delà des différences plus ou moins grandes qui séparent les langues du monde, on peut par un processus de comparaison reconstituer une sorte d'arbre généalogique qui permette de rendre compte des phénomènes de parenté typologique quand ceux-ci sont dus à une origine commune (et non au hasard ou à des phénomènes de contact entre les langues, comme c'est le cas des emprunts).

Ce ne sont pas toutes les familles de langues qui se prêtent aisément à cet exercice de reconstruction. En effet, l'existence d'une **abondante documentation historique** facilite grandement le travail des généalogistes de la langue, car elle permet de reconstituer avec plus de détails les différentes étapes évolutives subies par les strates successives de l'histoire d'une langue et de ses ramifications. Les conclusions que l'on peut tirer de la comparaison de ces données sont donc évidemment plus sûres. Or, il se trouve que dans l'ensemble des langues du monde, le groupe des langues romanes – auquel appartient le français – est particulièrement bien doté en documentation textuelle, historique, sur une période très longue : plus de deux millénaires, puisque les langues romanes remontent au latin.

En outre, ce groupe des langues romanes s'insère lui-même dans un ensemble plus large, celui des langues dites "indoeuropéennes". Nous allons dans les lignes qui suivent esquisser l'arbre généalogique de ces langues, en commençant par celles qui sont le plus près de nous, les langues romanes.

Les langues romanes.

On les définit simplement comme étant **les langues qui viennent du latin**, c'est-à-dire qui représentent l'aboutissement moderne de l'évolution du latin tel qu'on le parlait il y a plus ou moins deux mille ans, dans les différentes régions de l'Empire romain, et qui s'est transmis de génération en génération sans interruption jusqu'à nos jours. En fait, on peut dire que nous parlons tous encore latin aujourd'hui, nous qui parlons des langues romanes, mais un latin qui a subi deux mille ans d'évolutions et qui, surtout, s'est diversifié dans l'espace, pour donner naissance aux langues romanes, aussi appelées **langues néo-latines**.

Du point de vue démographique, on compte au premier rang de ces langues **l'espagnol**, parlé comme langue maternelle non seulement en Espagne mais aussi comme vous le savez dans plusieurs pays d'Amérique latine. Au deuxième rang, on a le **portugais**, qui occupe cette place (devant le français) car il est la langue maternelle non seulement du Portugal, mais aussi des habitants du Brésil, immense pays de près de 200 millions d'habitants ; cela dit, on parle également portugais dans certains pays africains. Vient ensuite le **français**, qui est la langue officielle d'un très grand nombre d'États, mais qui comme langue maternelle est tout de même beaucoup moins répandu. **L'italien** est la langue d'un pays relativement très peuplé, l'Italie, mais ce pays n'ayant pas connu de colonies outre-mer, le nombre de locuteurs d'italien est inférieur au nombre de locuteurs francophones. Cela dit, de nombreux immigrants italiens sont allés vivre ailleurs en Europe, ou en Amérique du Nord, en Amérique latine, voire en Australie, mais au bout de deux ou trois générations ils perdent l'usage de leur langue ancestrale et s'assimilent linguistiquement à leur pays d'accueil. Au cinquième rang des langues romanes, du point de vue du nombre de locuteurs, on retrouve le **roumain**, parlé en Roumanie (env. 21 millions d'hab.) et dans la Moldavie voisine (env. 4 millions d'hab.). Plusieurs personnes croient à tort que le roumain est une langue slave, méprise due au fait que cette langue a effectivement emprunté de nombreux mots aux langues slaves tout au long de son histoire, mais le fond du vocabulaire de cette langue, ainsi que son fonctionnement grammatical, sont indéniablement d'origine latine.

Ce sont les cinq langues romanes les plus importantes, mais nous allons voir la semaine prochaine qu'il y en a plusieurs autres, comme le **catalan** ou le **sarde**, ainsi que d'innombrables dialectes qui remontent au latin mais qui sont disparus entre temps, ou qui ne survivent qu'avec le statut social d'un patois.

Nous verrons dans le détail la semaine prochaine d'où viennent les langues romanes ; plus précisément, nous regarderons quelle était l'extension géographique du latin à l'époque de l'Empire romain, quelle sorte de latin s'est propagé dans l'Empire pour devenir l'ancêtre des actuelles langues romanes, comment ce latin s'est propagé, à quelle époque selon les régions, dans quelles conditions de transmission, etc., mais pour l'instant, nous allons d'abord resituer le groupe des langues romanes dans un ensemble plus vaste, à savoir la famille des **langues indoeuropéennes**.

De la même façon que l'on peut constater qu'il y a des similitudes, des ressemblances non fortuites entre les différentes langues romanes, on remarque aussi, en comparant certaines langues anciennes comme le **latin**, le **grec**, le **sanskrit**, le **gotique**, le **slave ancien**, qu'elles présentent entre elles des similitudes étonnantes, qui ne peuvent pas s'expliquer par le hasard. En fait, les linguistes se sont aperçus, depuis maintenant au moins deux siècles, que ces langues anciennes sont aussi apparentées les

unes avec les autres, et qu'elles remontent, en définitive, à un ancêtre commun. On appelle ces langues, ainsi que les langues modernes qui en dérivent, les “**langues indoeuropéennes**”, parce qu'elles étaient parlées à l'origine de l'Europe jusqu'à l'Inde (aujourd'hui, elles sont parlées sur les cinq continents, mais c'est une autre histoire). Quant à la langue dont elles dérivent toutes, on l'appelle, par convention, l'indoeuropéen commun, ou le proto-indoeuropéen, ou l'**indoeuropéen** tout court. Il faut remonter à **plus de 2000 ans avant J.-C.** pour retrouver l'époque où cette langue ancestrale n'avait pas encore commencé à se différencier en nombreux dialectes, qui allaient devenir à leur tour les langues indoeuropéennes anciennes. Il convient de préciser que ce que l'on en connaît n'est qu'une reconstruction, nécessairement incomplète, obtenue par la comparaison des langues indoeuropéennes anciennes.

Avant d'énumérer ces langues indoeuropéennes anciennes, ainsi que les langues modernes auxquelles elles ont donné naissance, il convient d'apporter une précision : ce ne sont pas toutes les langues traditionnellement parlées sur le sol de l'Europe qui font partie de la grande famille des langues indoeuropéennes. On compte trois exceptions : d'abord le **basque**, qu'on ne peut rattacher avec certitude à aucune autre famille de langues connue ; ensuite, ce qu'il est convenu d'appeler le groupe **finno-ougrien** (le hongrois d'une part ; le finnois, l'estonien et le lapon d'autre part) ; puis, comme une toute petite partie de la Turquie est située en Europe (au nord de la mer de Marmara), il faut aussi compter, parmi les langues parlées traditionnellement sur le sol de ce continent et n'appartenant pas aux langues indoeuropéennes, le **turc**. Donc, en résumé :

0. Langues non-indoeuropéennes d'Europe.

0.1. Le basque.

0.2. Le groupe finno-ougrien.

0.2.1. Le hongrois.

0.2.2. Le finnois (en Finlande).

0.2.3. L'estonien (en Estonie).

0.2.4. Le lapon (en Laponie, c'est-à-dire dans le nord de la Finlande).

0.3. Le turc.

Passons maintenant aux différents groupes de langues qui font partie de la grande famille indoeuropéenne. Certaines langues constituent un groupe à elles seules ; d'autres groupes réunissent de nombreuses langues. Voici une présentation par ordre alphabétique :

1. L'albanais.
2. L'arménien.
3. Les langues baltes.
4. Les langues celtiques.
5. Les langues germaniques.
6. Le grec.
7. L'indo-iranien.
8. Les langues romanes.

9. Les langues slaves.

10. Autres langues moins connues, aujourd'hui disparues.

Reprenons donc ces groupes un par un.

1. L'albanais.

Langue parlée bien sûr en Albanie, mais aussi dans quelques communautés au Kosovo, en Grèce et en Italie. Contrairement à ce que l'on pourrait être tenté de croire, l'albanais ne fait pas partie du groupe des langues slaves ; il constitue à lui tout seul un groupe à part entière au sein des langues indoeuropéennes.

2. L'arménien.

Langue parlée dans la République d'Arménie, petit pays du Proche-Orient qui faisait partie de l'ex-URSS, situé entre la Turquie, l'Iran, la Géorgie et l'Azerbaïdjan. Encore une fois, il s'agit d'une langue qui constitue un groupe à part entière au sein des langues indoeuropéennes.

3. Les langues baltes.

3.1. Le lituanien.

3.2. Le letton (ou *lette*).

3.3. Le vieux-prussien.

Ici, en revanche, nous avons affaire à un groupe qui connaît aujourd'hui deux représentants, à savoir le lituanien, parlé en Lituanie, et le letton (aussi appelé *lette*), parlé en Lettonie. Attention : des trois Républiques baltes, seules la Lituanie et la Lettonie parlent des langues baltes. L'Estonie, comme nous l'avons dit tout à l'heure, parle une langue, l'estonien, qui se rattache au groupe finno-ougrien, et qui n'a donc aucun rapport avec les langues indoeuropéennes. Faites surtout attention de ne jamais dire que le lituanien, l'estonien ou le letton sont des langues slaves ! Les Baltes ne vous le pardonneraient pas...

On ajoutera une troisième langue à ce groupe, celle appelée *vieux-prussien*, éteint depuis la fin du 17^e siècle par suite de la germanisation de la Prusse orientale.

4. Les langues celtiques (ou celtes).

4.1. Le celtique continental (langues aujourd'hui éteintes).

4.1.1. Le gaulois. Parlé en Gaule, est disparu dans l'Antiquité pour céder la place au latin.

4.1.2. Le celtibère. Parlé dans la Péninsule Ibérique, a aussi cédé la place au latin avant la fin de l'Antiquité.

4.2. Le celtique insulaire (langues aujourd'hui encore vivantes).

- 4.2.1. Le gaélique (en Irlande, puis en Ecosse).
- 4.2.2. Le brittonique (en Grande-Bretagne, puis en Bretagne).
- 4.2.3. Le gallois (Pays de Galles).
- 4.2.4. Le cornique (Cornouailles – avec un -s, ne pas confondre avec *la Cornouaille*, sans -s, qui est le nom d'une région de la Bretagne, en France).
- 4.2.5. Le breton (en Bretagne, France).

Il ressort de ce tableau que le gaulois n'est pas l'ancêtre direct du breton, bien qu'il fasse partie de la même famille. En fait, les ancêtres des Bretons sont arrivés en Bretagne aux 5^e, 6^e et début du 7^e s., chassés de Grande-Bretagne par les envahisseurs anglo-saxons, les ancêtres des actuels Anglais. Quand les ancêtres des actuels Bretons sont arrivés en Gaule, au début du Moyen Âge, les Gaulois – devenus entre temps des Galloromains – ne parlaient déjà presque plus le gaulois.

5. Les langues germaniques.

5.1. Le groupe oriental, dont la langue est connue sous le nom de *gotique* (attention : sans *h* pour la langue, avec *h* comme terme d'histoire de l'art et d'architecture). Toutes les variétés dialectales de cette langue se sont éteintes, les peuples qui les parlaient les ayant abandonnées pour le latin. On parle de groupe oriental car dans une première étape ils vivaient dans l'est de l'Europe, mais lors des grandes invasions germaniques ils se sont déplacés vers l'ouest, dans des territoires déjà latinisés :

- 5.1.1. Les Burgondes (c'est à eux que la Bourgogne doit son nom).
- 5.1.2. Les Wisigoths (Sud-Ouest de la Gaule et surtout Péninsule Ibérique).
- 5.1.3. Les Ostrogoths (qui, connus sous le nom de *Longobards* dans le nord de l'Italie, ont donné son nom à la *Lombardie*).
- 5.1.4. Les Vandales (le nom de l'Andalousie provient de leur nom).

5.2. Le groupe nordique (Scandinavie moins la Finlande).

- 5.2.1. L'islandais.
- 5.2.2. Le norvégien.
- 5.2.3. Le suédois.
- 5.2.4. Le danois.

5.3. Le groupe occidental.

- 5.3.1. Le francique (langue des Francs, envahisseurs germaniques qui conquièrent la Gaule, repoussant les Wisigoths vers l'Espagne et se mélangeant aux Galloromains, abandonnant leur langue au profit du latin tardif, non sans influencer celui-ci jusqu'à en faire l'ancêtre de l'actuelle langue française).
- 5.3.2. L'ancien haut-allemand (ancêtre de l'allemand standard moderne – *Hochdeutsch* – et de ses principaux dialectes, comme le bavarois et l'alémanique). Il faut également classer comme descendant de l'ancien haut-allemand le yiddish, l'ancienne

langue des Juifs ashkénazes d'Europe de l'Est, qui survit aujourd'hui dans certaines communautés juives en Europe, en Amérique et en Israël.

5.3.3. L'ancien bas-allemand (l'ancêtre du *Plattdütsch*, dialecte parlé dans les régions de l'Allemagne proches du Luxembourg et des Pays-Bas, mais aussi du néerlandais).

On parle en Afrique du Sud une langue appelée l'*afrikaans*, issue d'une variété de néerlandais exportée dans ce pays il y a deux siècles environ.

5.3.4. L'anglais.

6. Le grec.

Avec le grec, la situation est beaucoup plus simple qu'avec les langues germaniques ou romanes. A l'époque antique, il y avait en fait plusieurs dialectes, mais l'un d'entre eux s'est illustré davantage comme langue de culture : il s'agit de l'attique. À partir du 4^e siècle avant J.-C., cette variété, avec quelques simplifications (élimination de traits spécifiques), s'est imposée comme langue commune (*koinè*)¹ de tout le monde grec. Elle évoluera pour devenir le grec moderne, langue unitaire, contrairement au latin dont la répartition géographique trop étendue entraînera l'éclatement en plusieurs langues (v. ci-dessous 8).

7. L'indo-iranien.

Mais avant de passer aux langues romanes, voyons un autre groupe très important du point de vue démographique mais aussi pour des raisons d'historiographie scientifique, l'indo-iranien. On réunit les langues indiennes et les langues iraniennes en un seul groupe au sein des langues indoeuropéennes, parce qu'elles présentent plus de similitudes entre elles que ne le font les autres groupes de langues entre eux, ce qui suggère qu'à une époque très ancienne, elles ont dû être réunies, et avoir constitué une langue à part pendant une certaine période, avant de se scinder à nouveau en deux groupes.

7.1. Indo-aryen. Au 5^e s. avant J.-C., une variété littéraire de cet ensemble de dialectes est élaborée ; on la connaît sous le nom de *sanskrit*. C'est une des grandes langues de l'Antiquité, sur le même pied que le grec et le latin. C'est d'ailleurs l'étude des correspondances entre ces trois langues qui a donné son impulsion à la grammaire comparée, au 19^e siècle. À une certaine époque, on a cru que le sanskrit était l'ancêtre du grec et du latin, à cause de sa relative ancienneté, mais il n'en est que le cousin.

7.1.1. Hindi. Langue officielle en Inde, représentant moderne de l'ancien indo-aryen.

7.1.2. Urdu. Parlé dans le nord de l'Inde ainsi qu'au Pakistan, où il est langue officielle. En fait, le hindi et l'urdu sont presque la même langue (laquelle s'appellait autrefois *hindoustani*), mais l'urdu s'écrit en caractères arabes et le hindi en caractères sanskrits ('devanagari').

7.1.3. Tzigane. Les ancêtres des actuels tziganes sont partis du Nord-Ouest de l'Inde vers l'an 1000 de notre ère et se sont répartis peu à peu, çà et là en Europe, emportant leur langue avec eux.

¹ Le mot *koinésation* s'emploie d'ailleurs encore aujourd'hui en linguistique pour se référer à des phénomènes d'homogénéisation de conglomérats dialectaux.

7.2. Iranien. L'*avestique* est à l'iranien ce que le sanskrit est à l'indo-aryen, c'est-à-dire une ancienne forme codifiée de cette langue, dans laquelle plusieurs textes sacrés ont été rédigés.

7.2.1. Le persan (aussi appelé *farsi*), surtout en Iran (+ Afghanistan, Tadjikistan, etc.).

7.2.2. Le kurde.

7.2.3. Le pachtou (Afghanistan, Pakistan).

8. Les langues romanes.

Les langues romanes, nous l'avons dit, viennent du latin ; mais le latin faisait lui-même partie d'un groupe au sein des langues indoeuropéennes appelé "italique", puisque les langues en question se parlaient dans la Péninsule Italique. En plus du latin, qui s'est parlé d'abord dans la province du Latium, dont la capitale est Rome, la seule autre langue vraiment importante au sein du groupe italique est l'osco-ombrien ; mais elle disparaîtra, délogée par le latin, et non sans laisser quelques traces dans celui-ci. Quant au latin, la langue officielle de l'Empire Romain, il s'est étendu sur un territoire immense, et a donné lieu, par différenciation dialectale, à la grande famille des langues romanes, que nous allons étudier plus en détail la semaine prochaine. Nous allons nous contenter pour l'instant d'énumérer très grossièrement les différentes branches de cette famille, de l'ouest à l'est :

8.1. L'ibéroroman (la Péninsule Ibérique).

8.1.1. Le portugais (et sa langue-soeur, le gallego, pratiqué en Galice, province de l'extrême nord-ouest de l'Espagne, au nord du Portugal).

8.1.2. L'espagnol (variétés péninsulaires, variétés américaines ; judéo-espagnol des Sépharades).

8.1.3. Le catalan (de Catalogne, du Pays Valencien, des Îles Baléares).

8.2. Le galloroman (l'ancienne Gaule).

8.2.1. Le français et les patois d'oïl.

8.2.2. Les dialectes francoprovençaux (jamais unifiés, ne survivent qu'à l'état de patois).

8.2.3. Les dialectes occitans, ou la langue d'oc (le limousin, le languedocien, le provençal, ainsi que le gascon, à la personnalité très marquée, et qui pour cette raison est considéré par certains chercheurs comme une subdivision de la Galloromania à part entière).

8.3. L'italoroman (la Péninsule Italique).

Une certaine tradition regroupe sous cette appellation unitaire des sous-groupes de langues romanes essentiellement pour des raisons géo-politiques (grosso modo, leur appartenance à l'Italie ou leur grande proximité avec ce pays), mais à vrai dire les trois premiers (le rhétoroman, le sarde et le dalmate) devraient en fait être considérés comme des groupes à part entière au sein des langues romanes, plutôt que des dépendances de l'italoroman.

8.3.1. Le rhétoroman (le romanche en Suisse, le ladin dans les Dolomites et le frioulan dans le Frioul - ces deux derniers, dans l'extrême nord-est de l'Italie). Ce groupe lui-même existe davantage par commodité désignative que pour des raisons de proximité génétique ou typologique, car les trois groupes de parlers qui le constituent sont assez largement différents les uns des autres.

8.3.2. Le sarde (en Sardaigne). Du point de vue phonétique, l'une des plus archaïques des langues romanes.

8.3.3. Le dalmate (aujourd'hui éteint – le tout dernier locuteur est mort en 1898 ! –, se parlait autrefois en Dalmatie, région de la Croatie, sur la mer Adriatique).

8.3.4. L'italien à proprement parler (que l'on peut subdiviser à son tour en dialectes septentrionaux, toscans, et centre-méridionaux ; c'est le toscan qui est à la base de l'italien standard). On parle italien en Italie, en Suisse (surtout dans le canton du Tessin, mais aussi dans certaines vallées du canton des Grisons), et d'une certaine façon aussi en Corse (car les parlers traditionnels corses font partie de la grande famille des parlers italiens, et bien sûr pas de l'occitan ou du français).

8.4. Le balkanoroman (les Balkans).

8.4.1. Le dacoroumain (du nom de la Dacie, ancêtre de l'actuelle Roumanie). Parlé en Roumanie et en Moldavie.

8.4.2. Le macédonoroumain (parlé surtout en Macédoine).

8.4.3. Le méglénoroumain (nord-est de la Grèce).

8.4.4. L'istroroumain (parlé dans l'Istrie, une petite péninsule de la côte dalmate).

9. Les langues slaves.

Nous abordons maintenant le dernier grand groupe de langues indoeuropéennes, celui des langues slaves. Ce sont des langues assez étroitement apparentées, car à la fin du 8^e siècle après J.-C., elles ne s'étaient pas encore différenciées et ne représentaient qu'une seule et même langue. On peut identifier trois embranchements :

9.1. Le slave méridional. Ce sont, entre autres, les langues slaves de l'ex-Yougoslavie.

9.1.1. Le slovène.

9.1.2. Le serbe et le croate.

9.1.3. Le macédonien.

9.1.4. Le bulgare.

9.2. Le slave occidental. Ce sont les langues slaves de l'Europe centrale.

9.2.1. Le tchèque.

9.2.2. Le slovaque.

9.2.3. Le polonais.

9.3. Le slave oriental. Ce sont les langues slaves de l'Europe de l'Est.

9.3.1. Le russe.

9.3.2. Le biélorusse.

9.3.3. L'ukrainien.

On terminera avec quelques langues moins connues, aujourd'hui disparues.

10. Langues anciennes, moins connues et disparues.

10.1. L'anatolien (de *Anatolie*, nom qui désigne la partie de la Turquie située en Asie mineure, c'est-à-dire la plus grande partie de la Turquie), dont le principal représentant était le hittite, langue attestée par des documents qui datent du 17^e s. au 13^e siècle avant J.-C.

10.2. Le tokharien, groupe éteint depuis le 11^e siècle après J.-C. et qui se parlait dans le Turkestan chinois (extrême nord-ouest du pays) ; cette langue a la particularité d'être la langue indoeuropéenne ancienne la plus orientale de toute (alors que l'islandais est la plus occidentale de toutes).

Une dernière précision après avoir terminé ce tour d'horizon des langues indoeuropéennes : il ne faut surtout pas croire que parenté linguistique équivaut à parenté génétique ; en d'autres mots, tous les peuples, que ce soit dans l'Antiquité ou à l'époque moderne, qui parlent des langues indoeuropéennes, n'ont pas nécessairement les mêmes ancêtres immédiats ; les langues s'imposent le plus souvent par des conquêtes militaires, et se superposent à d'autres parlers, comme nous aurons souvent l'occasion de le voir dans le reste du semestre.

Lecture obligatoire cette semaine : l'encadré encyclopédique « L'indoeuropéen » (rédigé par G.-J. Pinault) dans le *Dictionnaire historique de la langue française* des éditions Le Robert (cet ouvrage est facile à trouver dans la section des ouvrages de référence de toute bibliothèque universitaire).